

Pädagogik

AFFEKTIVE WIRKUNGEN NORMORIENTierter PRÜFINSTRUMENTE IM FACH ENGLISCH AUF DIE EINSTELLUNG SEHR ÄNGSTLICHER UND WENIG ÄNGSTLICHER SCHÜLER

Eine experimentell angelegte Pilotstudie
im 5. und 6. Schuljahr

Inaugural - Dissertation
zur Erlangung des Doktorgrades

der

Philosophischen Fakultät

der

Westfälischen Wilhelmsuniversität zu
Münster (Westf.)

vorgelegt von

Kurt Haupt

aus Zossen/Brandenburg

(1977)

1978



Digitized by the Internet Archive
in 2026

Pädagogik

AFFEKTIVE WIRKUNGEN NORMORIENTierter PRÜFINSTRUMENTE IM FACH ENGLISCH AUF DIE EINSTELLUNG SEHR ÄNGSTLICHER UND WENIG ÄNGSTLICHER SCHÜLER

Eine experimentell angelegte Pilotstudie
im 5. und 6. Schuljahr

Inaugural - Dissertation
zur Erlangung des Doktorgrades

der

Philosophischen Fakultät

der

Westfälischen Wilhelmsuniversität zu
Münster (Westf.)

vorgelegt von
Kurt Haupt
aus Zossen/Brandenburg
(1977)

Tag d. mündl. Prüf. : 23. 10. 78
Dekan : Prof. Dr. M. E. Breckhoff
Referent : Prof. Dr. H. Blankertz
Korreferent : Prof. Dr. K. Sturze-
becher



D 6

INHALTSVERZEICHNIS

1. DIE GESELLSCHAFTLICHE BEDINGTHEIT VON PRÜFUNGEN UND PRÜFUNGSTECHNIKEN	1
1.1. Funktion von Prüfungen in der Waren produzierenden Gesellschaft	3
1.2. Die systemimmanente 'Adäquatheit' normorientierter Leistungsmessung	6
1.3. Kritik des Leistungsprinzips	9
2. WISSENSCHAFTSTHEORETISCHE PROBLEME EMPIRISCHER UNTERSUCHUNGEN	13
2.1. Methodologische Probleme	13
2.1.1. Das ungelöste Validitätsproblem	13
2.1.2. Relevanz positivistischer Forschungsdesigns vs. action research	15
2.1.3. Induktiv-statistische Systematisierungen	18
2.2. Probleme der Methodik	20
3. PROBLEMSTELLUNG	22
4. FORSCHUNGSDESIGN, -METHODEN, HYPOTHESEN UND STATISTISCHES INSTRUMENTARIUM. METHODISCHE UND METHODOLOGISCHE PROBLEME	24
4.1. Zum Begriff des pädagogischen Experiments	27
4.2. Untersuchungshypothesen	35
4.2.1. Geschlossene Hypothesen	40
4.2.2. Offene Hypothesen	43
4.3. Der Versuchsplan	44
4.4. Die experimentellen Gruppen	48
4.5. Das Untersuchungsinstrumentarium	54
4.5.1. Einstellungsskala 'Englisch'	57
4.5.1.1. Allgemeine methodologische Bemerkungen zur Einstellungsuntersuchung	58
4.5.1.2. Skalierungsmethode 'Paarvergleich'	59
4.5.1.3. Konstruktion der Einstellungsskala 'Englisch'	63
4.5.1.4. Evaluation der Einstellungsskala 'Englisch'	71

4.5.2. Das Semantische Differential	72
4.5.2.1. Methodische und methodologische Probleme	73
4.5.2.2. Das Rationale des Semantischen Differentials	76
4.5.3. Fragebogen für 'Prüfungsangst' und allgemeine Angst'	77
4.5.3.1. TASC-Itemanalyse	79
4.5.3.2. CMAS-Itemanalyse	80
4.5.3.3. Die Faktorenstruktur von CMAS und TASC	80
4.5.4. Fremdspracheneignungstest für die Unterstufe FTU 4-6	82
4.5.4.1. Exkurs (1): Grundannahmen der klassischen Testtheorie	83
4.5.4.2. Charakterisierung des FTU 4-6	84
4.5.5. Das Leistungs-Prüf-System (LPS) von Horn	85
4.5.6. Die informellen Tests I und II	87
4.5.6.1. Exkurs (2): Die Syntax informeller Tests und ihr Verhältnis zur schulischen Sozialisation und Identitätsfindung des Schülers	88
4.5.6.2. Beschreibungsskizze der informellen Tests I und II	97
5. UNTERSUCHUNGSERGEBNISSE	100
5.1. Stützung der These: Die Variation der Merkmalsausprägungen der abhängigen Variablen ist ein Effekt der experimentellen Bedingungen	101
5.2. Schüler des 5. Schuljahrs unterscheiden sich signifikant in der Beurteilung der Konzepte IX - XVI von den Schülern des 6. Schuljahrs	105
5.2.1. Einstufung des Konzepts 'SELBSTBILD'	108
5.2.2. Einstufung des Konzepts 'WUNSCHBILD'	110
5.2.3. Einstufung des Konzepts 'SCHWACHER SCHÜLER'	111
5.2.4. Einstufung des Konzepts 'GUTER SCHÜLER'	113
5.2.5. Einstufung des Konzepts 'ENGLISCHLEHRER'	114
5.2.6. Einstufung des Konzepts 'KLASSENARBEIT', 'ZENSUR' und 'KLASSENTEST'	115
5.3. Differentielle Analyse (1): Einstufungen der Konzepte IX - XVI durch wenig ängstliche und sehr ängstliche Schüler der Versuchsgruppe 5tes Schuljahr	116

5.3.1. Konzept ' <u>SELBSTBILD</u> ': Nicht die Art des Prüfinstruments, sondern der Grad der Ängstlichkeit determiniert das Selbstbild des Schülers auf den Dimensionen S und A	116
5.3.2. Konzept ' <u>WUNSCHBILD</u> ': Der informelle Test bewirkt Dissoziation: Die Diskrepanz zwischen 'Selbstbild' und 'Wunschbild' wird vergrößert	119
5.3.3. Konzept ' <u>SCHWACHER SCHÜLER</u> ': Der informelle Test hat einen dissoziativen Sozialisationseffekt: Der leistungsschwache Schüler wird nicht integriert	121
5.3.4. Konzept ' <u>GUTER SCHÜLER</u> ': Der Tendenz nach bewerten sehr ängstliche Schüler das Konzept auf der Dimension B positiver als wenig ängstliche	123
5.3.5. Konzept ' <u>ENGLISCHLEHRER</u> ': Weder die Prüfinstrumente noch der Grad der Ängstlichkeit determinieren signifikant die Einstufungen	124
5.3.6. Konzept ' <u>KLASSENARBEIT</u> ': Wenig ängstliche und sehr ängstliche Schüler zeigen nach der Bedingung 'informeller Test' auf der Dimension B sich signifikant unterscheidende Einstufungen	126
5.3.7. Konzept ' <u>ZENSUR</u> ': Das Stereotyp 'Zensur' wird weder durch die Prüfungsart noch durch den Ängstlichkeitsgrad verändert	128
5.3.8. Konzept ' <u>KLASSENTEST</u> ': Sehr ängstliche Schüler stufen den informellen Test auf der Dimension S signifikant positiver ein als wenig ängstliche	129
5.3.9. Zusammenfassung der Ergebnisse	131
5.4. Differentielle Analyse (2): Einstufungen der Konzepte IX - XVI durch wenig ängstliche und sehr ängstliche Schüler der Versuchsgruppe 6tes Schuljahr	135
5.4.1. Konzept ' <u>SELBSTBILD</u> ': Das Selbstbild sehr ängstlicher Schüler wird auf der Dimension S durch das Prüfinstrument determiniert	135
5.4.2. Konzept ' <u>WUNSCHBILD</u> ': Der informelle Test bewirkt Dissoziation: Persönlichkeitseigenschaften werden bevorzugt, die aggressives, konkurrentes Verhalten begünstigen	138
5.4.3. Konzept ' <u>SCHWACHER SCHÜLER</u> ': Die Einstufungen auf den Dimensionen B, S und A werden weder durch den Grad der Ängstlichkeit noch durch die Art des Prüfinstruments determiniert	140
5.4.4. Konzept ' <u>GUTER SCHÜLER</u> ': Die Prüfungsbedingungen lösen auf der Dimension B der Tendenz nach unterschiedliche Effekte aus	141
5.4.5. Das Konzept ' <u>ENGLISCHLEHRER</u> ': Die Bedingung 'informeller Test' bewirkt auf der Dimension S weniger positive Einstufungen als die Bedingung 'Klassenarbeit'	143

5.4.6. Konzept ' <u>KLASSENARBEIT</u> ': Nach der Klassenarbeit beurteilen sehr ängstliche Schüler im Gegensatz zu wenig ängstlichen das Konzept 'Klassenarbeit' auf der Dimension B signifikant negativer	144
5.4.7. Konzept ' <u>ZENSUR</u> ': Nach der Bedingung 'Klassenarbeit' beurteilen sehr ängstliche Schüler das Konzept auf der Dimension B negativer als wenig ängstliche	146
5.4.8. Konzept ' <u>KLASSENTEST</u> ': Wenig ängstliche Schüler bewerten den informellen Test nach der Bedingung 'Klassenarbeit' positiver als sehr ängstliche	148
5.4.9. Zusammenfassung der Ergebnisse	150
5.5. Die Prüfungsbedingungen bewirken keine Einstellungsänderungen bei wenig ängstlichen oder sehr ängstlichen Schülern	153
6. ZUSAMMENFASSUNG	155
ANHANG	162
1. Informeller Test I (5tes Schuljahr)	162
1.1. Anweisungen für die Prüfung (Tonbandtext)	162
1.2. Testbogen 'INFORMELLER TEST I (5)'	165
2. Informeller Test II (6tes Schuljahr)	172
3. Einstellungsskala 'Englisch'	176
4. Angstskalen	178
5. Meßinstrument 'Semantisches Differential'	181
6. Verzeichnis der Tabellen	184
7. Verzeichnis der Abbildungen	186
8. Verzeichnis der Grafiken	187
LITERATURVERZEICHNIS	188

1. DIE GESELLSCHAFTLICHE BEDINGTHEIT VON PRÜFUNGEN UND PRÜFUNGSTECHNIKEN

Kritische Erziehungswissenschaft ist ihrem Selbstverständnis nach Gesellschaftswissenschaft. Die wissenschaftstheoretische Reflexion des Entstehungs-, Begründungs- und Verwertungszusammenhangs von Wissenschaft spannt die Koordinaten auf, die für die eigene Standortbestimmung wesentlich sind. Für den empirisch arbeitenden Erziehungswissenschaftler ergibt sich daraus die Konsequenz, das kritische Potential wissenschaftstheoretischer Analysen z.B. bei der Auswahl des Forschungsproblems und der Methode, bei der Operationalisierung eines Konstrukts usw. zu berücksichtigen, d.h. forschungsprogrammatisch "unverkürzte Rezeption der empirisch-analytischen Verfahren bei gleichzeitiger ideologiekritischer Hinterfragung ihrer politisch-gesellschaftlichen Funktion, und d.h. zugleich die hermeneutische Identifikation des leitenden Erkenntnisinteresses bei der Erforschung edukativer Vorgänge im unerläßlichen Kontext mit der allgemeinen wissenschaftstheoretischen Diskussion in den Sozialwissenschaften, aber erarbeitet aus der Eigenstruktur der Erziehung."¹⁾

Dabei sollte der Empiriker nicht falsche Erwartungen an wissenschaftstheoretische Aussagen knüpfen, weil "Wissenschaftstheorie eine R e f l e x i o n bestehender oder zumindest konzipierter Forschungspraxis (ist). Wissenschaftstheorie hat nicht instrumentalen, sondern reflexiven Charakter."²⁾

Die nicht zu unterschätzende Leistung der Methoden empirischer Sozialforschung, funktionale Zusammenhänge symbolisch zu repräsentieren, erfährt ihre Relativierung durch die gesellschaftskritische Funktion wissenschaftstheoretischer Begriffsanstrengung. Exemplarisch dafür mag als Beispiel die prognostische

-
- 1) BLANKERTZ, H.: Pädagogik unter wissenschaftstheoretischer Kritik, in: Oppolzer, S. (Hrsg.), unter Mitwirkung von R. Lassahn: Erziehungswissenschaft 1971 zwischen Herkunft und Zukunft der Gesellschaft, Wuppertal 1971, S. 31
 - 2) MEYER, H.L.: Trainingsprogramm zur Lernzielanalyse, Frankfurt am Main 1974, S. 18

Validität von Zensuren für die individuelle Prognose des späteren Berufserfolgs stehen¹⁾):

- a) Die Methoden der empirischen Sozialforschung erlauben es, einen Koeffizienten zu berechnen, der den Zusammenhang von Prädiktor und Kriterium quantifiziert.
- b) Die wissenschaftstheoretische Analyse²⁾ deckt jedoch die unbemerkt gebliebenen Voraussetzungen auf, die Annahme nomologischer Gesetzmäßigkeiten nach dem H - O - Schema³⁾. Solch eine Invarianz existiert aber nicht im Prognosebereich, so daß die Prognose eines individuellen Falles sehr fehlerhaft ist.
- c) Die gesellschaftskritische Reflexion betrifft die Frage, wie es kommt, daß trotz der geringen prognostischen Validität von Zensuren diese beim Numerus clausus eine so große Rolle spielen, "daß Zensuren (obwohl sie) einerseits ein untaugliches Mittel darstellen, andererseits trotzdem weiterhin Verwendung finden."⁴⁾

Ziehen wir das Fazit dieser grundsätzlichen Überlegungen: Eine Legitimierung vorherrschender Prüfungs- und Berechtigungspraktiken durch ihr bloßes faktisches Vorhandensein reicht nicht aus; ebenso ein Erklärungsschema wie das HEMPEL-OPPENHEIMERSche für die Erklärung individueller Tatbestände. Ebenso kritisch muß der Absolutheitsanspruch einer Theorie hinterfragt werden, die für wahr hält, "was in all diesen Fällen doch immer neu zu erweisen ist: daß das Substrat des Geschichtsverlaufs und also aller gesellschaftlichen Praxis einerseits die materielle Pro-

1) vgl. KVALE, St.: Prüfung und Herrschaft. Hochschulprüfungen zwischen Ritual und Rationalisierung, Weinheim/Basel 1972, S. 49 ff.

2) KROPE, P.: Zensur, Ausbildung und Gesellschaft. Zur Kritik und Alternative eines Bewertungsverfahrens, Essen 1976, S. 22 ff.

3) STEGMÜLLER, W.: Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band I. Wissenschaftliche Erklärung und Begründung. Studienausgabe, Teil 1, Berlin/Heidelberg/New York 1969, S. 86

4) KROPE, P., a.a.O., S. 58

duktion ist und andererseits die Gesamtheit der pädagogischen 'Erscheinungen' auch wirklich erreicht."¹⁾

Welchen Stellenwert der zuletzt genannte Ansatz für unsere Problemstellung hat, soll nun unter anderem erörtert werden.

1.1. Funktion von Prüfungen in der Waren produzierenden Gesellschaft

In seiner Monografie weist DOHSE nach, daß das Entstehen des Phänomens 'Schulzeugnis' hauptsächlich - wenn auch nicht restlos - mit der Veränderung gesellschaftlicher Strukturen erklärt werden kann. "Wichtigstes Ergebnis ist die Erkenntnis, daß das Schulzeugnis seinem Wesen nach Überwiegend soziologisch zu begreifen ist. In seiner traditionellen Gestalt hat es sich zu einem Hilfsmittel einer bürokratisierten, nationalstaatlich organisierten Gesellschaft im Dienste der Heranbildung und Auslese ihres Nachwuchses auf der Grundlage des Leistungsprinzips entwickelt."²⁾ Dies bedeutete auf dieser historischen Stufe einen Fortschritt in Richtung sozialer Chancengleichheit, da nun nicht mehr die soziale Herkunft den sozialen Aufstieg behinderte oder begünstigte, sondern die "persönliche Tüchtigkeit zum Auslesekriterium"³⁾ avancierte, so daß - zumindest formal - das Gewicht ständischer Privilegien verringert wurde.

In der neueren Diskussion über Zensuren und Prüfungen konstituierten sich zwei Positionen, die aufgrund ihrer unterschiedlichen Fragestellungen zum gleichen Untersuchungsgegenstand zwei divergierende Wissenschaftstheorien konturieren. INGENKAMPS Reader "Die Fragwürdigkeit der Zensurengebung"⁴⁾ enthält unaus-

1) MOLLENHAUER, K.: Theorien zum Erziehungsprozeß. Zur Einführung in erziehungswissenschaftliche Fragestellungen, München 1972, S. 175

vgl. dazu auch NYSSSEN, F. (Hrsg.): Schulkritik als Kapitalismuskritik, Göttingen 1971, S. 25; zitiert nach: NEUMANN, G.: Arbeit, in: WULF, Chr. (Hrsg.): Wörterbuch der Erziehung, München/Zürich 1974, S. 30

2) DOHSE, W.: Das Schulzeugnis. Sein Wesen und seine Problematik, Weinheim 1963, S. 126

3) DOHSE, W., a.a.O., S. 17

4) INGENKAMP, K.: Die Fragwürdigkeit der Zensurengebung, Weinheim 1971²

gesprochen die Auffassung des logischen Empirismus. Im Vordergrund steht die methodisch-statistische Analyse, deren Ziel die Übertragung testtheoretischer Gütekriterien auf die herrschende Prüfungspraxis ist, so daß methodologische Fragestellungen durch die Beschränkung auf prüfungsimmanente Probleme nicht thematisiert werden. Eine gegensätzliche, die gesellschaftlichen Rahmenbedingungen mit einbeziehende Position vertritt KVALE. Seiner Meinung nach reflektieren die Widersprüche im Prüfungswesen die Antinomien der Gesellschaft. Prüfungen dienen der Stabilisierung von Herrschaftsverhältnissen. Eine Lösung dieser Widersprüche bringe erst die Veränderung der Produktionsbedingungen: "Denn allenfalls die radikale Umwälzung der Herrschafts- und Eigentumsverhältnisse einer Klassengesellschaft könnte das partikulare System der Prüfungen materiell verändern."¹⁾

Beide Ansätze reichen unserer Meinung nach nicht aus für eine Humanisierung des Prüfungs- und Zensurenwesens. Während der erste Ansatz die Frage nach Sinn und Inhalt von Leistungen ausklammert, ihre humanen oder inhumanen Wirkungen innerhalb eines Gesellschaftssystems nicht untersucht, suggeriert die zweite Art des Zugriffs eine Problemlösung via Veränderung der Produktionsbedingungen; die man in den sozialistischen Staaten findet. Ge-rechterweise muß nachgetragen werden, daß KVALE zugibt, daß z.B. die DDR, ein sozialistischer Staat, sich in ihrem Prüfungswesen kaum von einem kapitalistischen System unterscheidet.²⁾ Die Klärung dieses Widerspruchs durch die Attributierung 'revisio-nistischer' oder 'revolutionärer' Sozialismus³⁾ ist jedoch un-befriedigend und auch zu vage. Wir sind der Meinung, daß über die Änderung der Produktionsbedingungen hinaus ein rationaler Diskurs über die Legitimation der Lern- und Leistungsziele erst hinreichend ist. Die politische Komponente hebt DREITZEL hervor: "Da aber eben Zweckrationalität des Verhaltens in bestimmten Handlungsbereichen ein funktionaler Imperativ industriell pro-

1) KVALE, St., a.a.O., S. 240

2) KVALE, St., a.a.O., S. 225

3) KVALE, St., a.a.O., S. 227

duzierender und bürokratisch verfaßter Gesellschaften bleibt, beschreibt die gesellschaftliche Dissoziation von Lust und Leistung einen Grundwiderspruch der gegenwärtigen Produktionsverhältnisse, der für die sozialistischen Gesellschaften solange ebenso bestimmend bleibt wie für die kapitalistischen, wie diese das Potential veränderter Eigentumsbedingungen nicht durch eine fundamentale Demokratisierung im politischen und administrativen Bereich realisieren können."¹⁾

Wenn wir einen Zusammenhang zwischen Prüfungssystem und Gesellschaft²⁾ postulieren, dann gehen wir von zwei Voraussetzungen aus:

1. Individuum und Gesellschaft bilden ein dialektisches Verhältnis.
2. Die materiellen Verhältnisse konstituieren gesellschaftliche Wirklichkeit.

Für die BRD gilt eine kapitalistische Produktionsweise mit den Merkmalen: Privateigentum an den Produktionsbedingungen, Arbeitsteilung, Austausch der Arbeitsprodukte. Zentrale ökonomische Kategorie ist die der Ware, die sowohl einen Gebrauchs- als auch einen Tauschwert besitzt. Waren werden produziert, um Mehrwert zu schaffen. Mehrwertvergrößerung wird jedoch nur dann weiterhin erzielt, wenn qualifizierte Arbeitskräfte ausgebildet werden. Das Geschehen im Bildungssektor hat deshalb wesentliche Bedeutung für die Erhaltung des Gesamtsystems. Vom analytischen Standpunkt aus gesehen hat das Prüfungssystem zwei Funktionen zu erfüllen, "nämlich erstens die Qualifikation für und zweitens die ideologische Integration in den künftigen Tätigkeitsbereich."³⁾

Die erworbenen Qualifikationen werden in der Regel durch die Zensur repräsentiert. Paradoxerweise aber gibt die Ziffernzensur keinen inhaltlichen Hinweis auf die dahinter stehenden Kenntnisse und Fertigkeiten. "Ihre Abstraktionsform spiegelt die

1) DREITZEL, H.P.: Soziologische Reflexionen über das Elend des Leistungsprinzips, in: Sinn und Unsinn des Leistungsprinzips. Ein Symposium, München 1974, S. 31

2) KROPE, P., a.a.O., S. 59 ff.

3) KROPE, P., a.a.O., S. 73

Abstraktionsform des auf Warenproduktion und -tausch beruhenden Gesellschaftssystems, in dem der Tauschwert einer Ware zum beherrschenden Motiv und Zweck wird. Die Zensurengebung spiegelt die Vorherrschaft der Tauschwerte über die Gebrauchswerte in der Waren produzierenden Gesellschaft wider ... Zu dieser zweiten Funktion, der ideologischen Integration in den Arbeitsprozeß, kann die Zensurengebung beitragen, indem sie durch die Identität der Abstraktionsform ihres methodischen Ansatzes mit der des Arbeitsprozesses dessen Ideologie in den Ausbildungssektor vorverlagert. Die Zensurengebung erzeugt die Ziele und Inhalte der Ausbildung nicht, aber sie begünstigt und verstärkt sie durch ihren Selektionsmechanismus."¹⁾

1.2. Die systemimmanente 'Adäquatheit' normorientierter Leistungsmessung

Unsere Generalhypothese lautet: Der normorientierte Bewertungsansatz entspricht "der einseitigen Orientierung .. an der Funktion der Ausbildung als Zulieferer für den Abnehmerbereich."²⁾

Der Mittelwert der Bezugspopulation fungiert als Orientierungspunkt bei der vergabe 'guter' und 'schlechter' Noten, und er erlaubt die rangmäßige Einstufung der Schüler. Die Messung der Schülerleistung zielt also nicht darauf festzustellen, wieweit die Sache beherrscht wird, das Lernziel erreicht worden ist, sondern dient letzten Endes der Zuordnung eines Rangplatzes für Selektionszwecke. Dem Selektionsmechanismus korrespondiert auf der Schülerseite die Erfahrung, daß der Vorrat guter Noten knapp ist. An dieser Stelle wird der Zusammenhang des ökonomischen Bereichs mit dem Bildungsbereich deutlich. "Wenn Ausbildungsplätze und qualifizierte Arbeitsmöglichkeiten knapp sind, ermöglicht der Rangplatz eines Lernenden die Zuweisung seines weiteren Lernweges und seiner beruflichen Position. In dem Maße, in dem Knappheit aber aus dem ökonomischen Kalkül des Produktionssektors resultiert, stellt das hierarchische Verfahren der Zensurenbewertung einen Vorgriff ~~des~~ hierarchisch arbeitsteiligen Produkti-

1) KROPE, P., a.a.O., S. 79 - 80

2) KROPE, P., a.a.O., S. 76

onsprozesses auf die Ausbildung dar."¹⁾

Die starke Betonung der selektiven Funktion bei der normorientierten Leistungsmessung geht zu Lasten der diagnostischen Funktion oder eines lernzielorientierten Ansatzes, der sich nicht an der Knappheit von Ausbildungsplätzen - siehe Numerus clausus - orientiert, sondern den Schülern Auskunft darüber gibt, "wie gut jeder von ihnen in Bezug auf jene Aufgaben ist, die die Operationalisierung des Bildungsrechts darstellen könnten"²⁾.

Normorientierte Leistungsmessung und ihr Zuteilungsmechanismus von Zensuren erfüllen in der Ausbildung vier Funktionen: Selektion, Arbeitsintensivierung, Konformitätsdruck und ideologische Integration.³⁾ Eine fünfte Funktion, die Lernzielorientierung, kann sie jedoch in der Ausbildung für die Waren produzierende Gesellschaft nicht leisten, nämlich den "Nachweis eines geforderten durchschnittlichen Qualifikationsniveaus."⁴⁾ Jetzt wird ihr methodisch-statistischer Ansatz zum Problem. Die vorgeschlagenen Lösungsversuche konservieren die vier genannten Funktionen und integrieren die fünfte durch Kombination dreier Elemente: lernzielbezogene Leistungsbewertung, Bezug auf die klassische Testtheorie und Zuordnung der Testpunkte zu Noten. Der Legitimation von Lernzielen wird dabei jedoch keine Aufmerksamkeit geschenkt.

Das Prüfungsinstrument ist zwar geeignet, "individuelle Unterschiede hinsichtlich des gemeinsamen Leistungsstandards " zu erfassen, doch ungeeignet für die Aufnahme "gemeinsame(r) Leistungen"⁵⁾. Solidarisches Verhalten oder das "Einüben einfühlenden und liebevollen Handelns, das den Partner nicht länger in marktkonformen Kategorien des Tauschwertes taxiert"⁶⁾

1) KROPE, P., a.a.O., S. 75

2) KROPE, P., a.a.O., S. 75

3) KROPE, P., a.a.O., S. 80

4) KROPE, P., a.a.O., S. 80

5) ULICH, D.: Gruppendynamik in der Schulklasse. Möglichkeiten und Grenzen sozialwissenschaftlicher Analysen, München 1971, S. 63

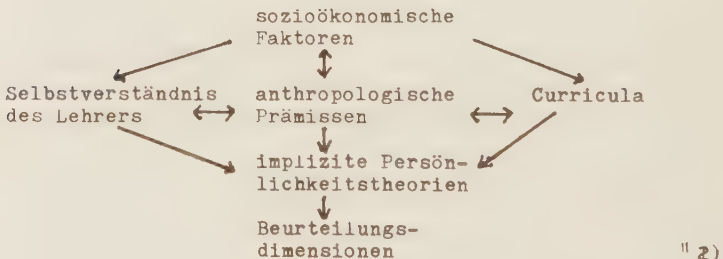
6) ULICH, D./MERTENS, W.: Urteile über Schüler. Zur Sozialpsychologie pädagogischer Diagnostik, Weinheim/Basel 1973, S. 160

sind Leerstellen in der Matrix dieses Kalküls.

Objektivisierte Leistungsmessung, die sich nur auf die Verbesserung methodenimmanenter Standards konzentriert, den gesellschaftlichen Rahmen jedoch unreflektiert läßt, liefert die Objekte ihrer 'Behandlungen' den herrschenden Interessen aus, macht sie ihnen verfügbar. Eine alternative bietet eine kritische Sozialforschung, welche die Individuen zum Subjekt der Forschung macht, und der es "gelingt, 'Erklärung' zu 'Aufklärung' zu machen: Erkenntnis in Bewußtheit zu überführen. Die 'abhängigen Variablen' der Sozialforschung, die Individuen selbst, müssen zu 'unabhängigen Variablen' werden: Durch Erklärung aufgeklärt verändern sie möglicherweise die Bedingungen, durch die sie zu 'Abhängigen' gemacht wurden."¹⁾

Eine Verbesserung der Objektivität, Reliabilität und Validität von Prüfinstrumenten lenkt nur davon ab, "die den impliziten Persönlichkeitstheorien vorgeordneten Faktoren zu revidieren. Bei einem solchen Versuch wird aber auch deutlich, daß einer Revision Grenzen gesetzt sind, weil die als antezedierend bezeichneten variablen in Interdependenz mit sozioökonomischen Faktoren stehen. Der Zusammenhang sei in einem Schema noch einmal verdeutlicht:

Der Bedingungs Zusammenhang von Beurteilungsdimensionen



1) Ulich, D. / Mertens, W., a.a.O., S. 13

2) Ulich, D. / Mertens, W., a.a.O., S. 128 - 129

1.3. Kritik des Leistungsprinzips

"Der Begriff der Leistung ist mit so vielen Äquivokationen behaftet, daß sich seine wissenschaftliche Verwendung fast verbietet"¹⁾, und "vor allem die affektive Belastetheit des Leistungsbegriffs läßt ihn als wissenschaftlichen Terminus wenig geeignet erscheinen."²⁾

Im schulischen Bereich hat der Begriff Leistung nach FURCK vier Bedeutungen: "Leistung als schulische Forderung an den Schüler, Leistung als Tätigkeit des einzelnen innerhalb der verschiedenen Leistungsbereiche und Leistung als besonderer Beitrag der Schule für Gesellschaft, Staat, Wirtschaft und Wissenschaft. Diese formale Beschreibung sagt über den jeweils spezifischen Charakter der Leistungsforderung oder der Leistung und über ihre erzieherische Bedeutung noch nichts aus, hierüber entscheidet erst ihre inhaltliche Bestimmung."³⁾ Die Inhalte widerspiegeln jedoch Leistungsansprüche außerpädagogischer Instanzen zum Zwecke ihrer Interessendurchsetzung. "Für die pädagogische Lösung des Leistungsproblems ist zwar die Frage nach dem angemessenen Verhältnis von objektivem Leistungsanspruch und individueller Leistungsfähigkeit wichtig, wenn Erziehung jedoch ihren Zweck in der Emanzipation des Menschen hat, so ist die Erziehung zum rationalen und kritischen Hinterfragen der gegebenen Ansprüche auf ihre individuelle und gesellschaftliche Relevanz die *conditio sine qua non*. Erziehung zur Leistungsbereitschaft kann demzufolge immer nur heißen Erziehung zur k r i t i s c h e n Leistungsbereitschaft. D.h. zu einer Haltung, die nicht nur sich selbst etwas abverlangt, sondern weiß, w a r u m u n d z u w e l c h e m Z w e c k e i n e L e i s t u n g a n g e s t r e b t w i r d. Es wird zu prüfen sein, wie weit und auf welchem Niveau die Schüler in den verschiedenen Altersstufen zu derartigen Re-

-
- 1) DREITZEL, H.P.: Soziologische Reflexionen über das Elend des Leistungsprinzips, in: Sinn und Unsinn des Leistungsprinzips. Ein Symposium, München 1974, S. 31
 - 2) KRUMM, V.: Leistung - Schulleistung - Schulleistungsmessung, in: WULF, Chr. (Hrsg.): Wörterbuch der Erziehung, München / Zürich 1974, S. 383
 - 3) FURCK, K.-L.: Das pädagogische Problem der Leistung in der Schule, Weinheim 1961, S. 118

flexionen angeleitet werden können."¹⁾ Den größten Schulerfolg scheint jedoch der Schüler zu haben, "der sich an die rational nicht begründbaren Leistungsnormen am besten anpaßt, ohne nach Sinn und Zweck dieser Leistungen zu fragen."²⁾ Dem 'Erfolg' entspricht die mangelhaft ausgebildete Fähigkeit zum kommunikativen Handeln. "Das schulische Leistungssyndrom, das sich in ritualisierten und formalisierten Prüfungen manifestiert, fördert den instrumentellen (analytischen) Sprachgebrauch zuungunsten des reflexiven. Diese I n s t r u m e n t a l i s i e r u n g äußert sich in der distanzlosen Übernahme handlungsbezogenen Wissens der Gesellschaft als Lehr- und Lernstoff."³⁾

Die systemstabilisierende und Chancengleichheit legitimierende ideologische Funktion des Leistungsprinzips demonstriert die folgende Argumentationsfigur⁴⁾:

1. Jeder hat die gleichen Möglichkeiten aufzusteigen.
2. Der soziale Rang des einzelnen in der Gesellschaft ist das Ergebnis freien Wettbewerbs. Doch kann gezeigt werden, "daß Berufsrollen nach einem Syndrom von Kriterien vergeben werden, das nur zu einem geringen Teil aus Leistungskriterien besteht. Die industrielle Gesellschaft ist weit davon entfernt, eine Leistungsgesellschaft zu sein. Für den Einzelnen sind die Möglichkeiten, seinen Leistungen entsprechend voranzukommen oder sozial abzustiegen, sehr eingeschränkt."⁵⁾
3. Jeder ist für seinen Erfolg oder Mißerfolg selbst verantwortlich.
4. Diejenigen, welche den höchsten Rangplatz innehaben, haben dafür auch die höchsten Leistungen erbracht; die weniger Erfolgreichen waren faul, ohne Ehrgeiz, oder es fehlte ihnen an der notwendigen Begabung.

1) NEUHAUS, E.: Zum pädagogischen Leistungsbegriff, in: LICHTENSTEIN- ROTHER, I. (Hrsg.): Schulleistung und Leistungsschule, Bad Heilbrunn/Obb. 1971, S. 11

2) RAUCH, E./ANZINGER, W. (Hrsg.): Wörterbuch Kritische Erziehung, Starnberg 1972, S. 271

3) HEINZE, Th.: Zur Kritik an den Technologisierungstendenzen des Unterrichtsprozesses, in: Die Deutsche Schule, 1972, S. 360

4) Wir halten uns dabei an: SEIBEL, H.D.: Gesellschaft im Leistungskonflikt, Düsseldorf 1973, S. 59 ff.

5) SEIBEL, H.D., a.a.O., S. 57

5. So entsteht eine optimale gesellschaftliche Ordnung. "Jede Abweichung ...stellt eine Verletzung des 'gerechten' Leistungsprinzips dar.

So können mit der Leistungsideologie alle Privilegien und alle Arten von Ausbeutung, kurz: jede Klassenlage legitimiert werden. Damit wird dem Leistungsprinzip von seinen Apologeten nicht nur normative, sondern auch faktisch-deskriptive Bedeutung zugesprochen: Unsere Gesellschaft sollte nach den Leistungsapologeten nicht nur danach streben, eine Leistungsgesellschaft zu sein; sie ist es tatsächlich."¹⁾

Da nun die Statuszuweisung nicht aufgrund von Kriterien erfolgt, die das Ergebnis einer umfassenden Analyse des Arbeitsplatzes sind, wird das postulierte Leistungsprinzip destruiert; es kommt für den einzelnen, der Erfolg haben will, nur darauf an, sich mit den herrschenden Normen zu identifizieren.²⁾ Auf schulische Lernprozesse übertragen heißt das, daß der Anpassungszwang, der von fremdbestimmtem Wissen ausgeht, dem gesellschaftliche Wertmaßstäbe entsprechen, bestimmte Dispositionen präferiert. "Die Leistungsideologie der Schule macht dumm, weil sie die differenzierteren Individuationsprozesse im Schüler nicht fördern kann, die von den gangbaren Schneisen abweichen. Leistungsbereitschaft ohne kritische Sozialtheorie trägt aber immer schon faschistische Züge."³⁾ "Wer im schulisch unterstützten Vorgang seiner Individuation nicht zur Entfaltung seiner wesentlichen Dispositionen geführt wird, kann es kaum nachholen, selbst wenn er durch glänzende Entlohnung seiner Leistung honoriert wird. Durch die schulische Einengung auf Leistung ist er dann das verkümmerte, aber disponible Produkt allmächtiger Apparate, denen er ausgeliefert ist. Gerade sein hervorragendes Leistungsvermögen kann ihn also daran hindern, seine humanen Möglichkeiten wahrzunehmen. Er ist um sein Menschsein betrogen. Dieser schulische Prozeß entspricht präzise dem von K a r l M a r x vielfältig beschriebenen Entfremdungsvorgang im Kapitalismus und bestimmt auch die kapita-

1) SEIBEL, H.D., a.a.O., S. 60

2) vgl. OFFE, C.: Leistungsprinzip und industrielle Arbeit, Frankfurt am Main 1970, S.33, sinngemäß zitiert nach: MOSER, H.: Leistungsprinzip: Ökonomische Kategorie und soziale Norm, in: Die Deutsche Schule, 1972/3, S. 206

3) GAMM, H.-J.: Kritische Schule. Eine Streitschrift für die Emanzipation von Lehrern und Schülern, München 1970, S.117-118

listische Schule."¹⁾

Es ist daher zu vermuten, daß als Folge der Erziehung zu entfremdeter Leistung, d.h. zu "Leistungen, die dem Individuum von außen auferlegt werden und deren Zusammenhang im Produktionsprozeß es nicht durchschaut"²⁾, eine Persönlichkeit entsteht, die ein instrumentales Verhältnis zum Lernen entwickelt. Die Frage nach dem Sinn und Inhalt einer Leistung verschwindet aus ihrem Denkhorizont. An die Stelle der konkreten Leistung tritt die abstrakte. Strukturmerkmale der Gesellschaft affizieren - wenn auch nicht total - Persönlichkeitsstrukturen. Symptom dafür ist die Dissoziation von Lust und Leistung. "Da aber eben Zweckrationalität des Verhaltens in bestimmten Handlungsbereichen ein funktionaler Imperativ industriell produzierender und bürokratisch verfaßter Gesellschaften bleibt, beschreibt die gesellschaftliche Dissoziation von Lust und Leistung einen Grundwiderspruch der gegenwärtigen Produktionsverhältnisse, der für die sozialistischen Gesellschaften solange ebenso bestimmend bleibt wie für die kapitalistischen, wie diese das Potential veränderter Eigentumsbedingungen nicht durch eine fundamentale Demokratisierung im politischen und administrativen Bereich realisieren können"³⁾.

"Wenn es auch nicht Leistung als solche abzuschaffen gilt, so geht es jedoch darum, die Entfremdung der Leistung aufzuheben und - als ersten Schritt - die/Ideologie der Leistungsgesellschaft zu entlarven, die ein afunktionales Leistungsprinzip dazu benützt, die Menschen an den Status quo der spätkapitalistischen Industriegesellschaft anzupassen. In dieser Beziehung scheint bemerkenswert, daß im pädagogischen Bereich eine Bedeutung des Begriffes 'Leistung' aufbewahrt wird, die Leistung und Spiel miteinander in Beziehung bringt und so die Erinnerung an ein nicht entfremdetes Leistungsprinzip beinhaltet."⁴⁾

1) GAMM, H.-J., a.a.O., S. 116 - 117

2) MOSER, H.: Leistungsprinzip: Ökonomische Kategorie und soziale Norm, in: Die Deutsche Schule, 1972, S. 209

3) DREITZEL, H.P., a.a.O., S. 50

4) MOSER, H., a.a.O., S. 211

2. WISSENSCHAFTSTHEORETISCHE PROBLEME EMPIRISCHER UNTERSUCHUNGEN

Terminologisch kann die *M e t h o d i k* einer Wissenschaft unterschieden werden von der "Theorie der wissenschaftlichen Methoden oder der Methoden, die prinzipiell ~~von~~ bzw. in jeder Wissenschaft angewendet werden können"¹⁾, der *M e t h o d o l o g i e*. Für unseren Problemzusammenhang kommen nur einige spezielle Probleme empirischer Untersuchungen in Frage. Von der weiteren Differenzierung des Begriffs Methodik in eine Objekt- und Metamethodik ist für unsere Zwecke der Terminus Objektmethodik von Bedeutung, da er "das System von Methoden (bezeichnet), das diese Wissenschaft auf ihren Gegenstand anwendet. Zwischen den Objektmethodiken verschiedener Wissenschaften können ... relativ große Unterschiede bestehen, da die Anwendbarkeit einer Methode wesentlich durch den Gegenstand einer Wissenschaft und durch deren spezifische Weise der Erforschung dieses Gegenstandes mitbestimmt wird."²⁾ Doch sind Methodik und spezielle Methodologie einer Wissenschaft keine synonymen Begriffe.

2.1. Methodologische Probleme

2.1.1. Das ungelöste Validitätsproblem

Die Theoriebildung in den Sozialwissenschaften ist auf Begriffe angewiesen, die man als 'hypothetische Konstrukte' bezeichnet. Sie stehen stellvertretend für Dimensionen, die nicht direkt beobachtbar sind.

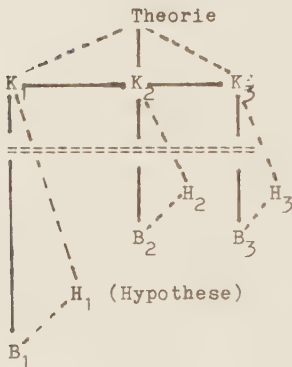
In der vorliegenden Arbeit benutzen wir z.B. den theoretischen Begriff 'Angst', der einen gewissen Bedeutungsüberschuß besitzt, der darin besteht, "daß der Wissenschaftler über das 'Ängstlichkeit' Genannte immer wieder Aussagen zu machen hat, die noch nicht mit Hilfe von entsprechenden Zuordnungsregeln auf Beobachtungen 'reduziert' sind und für welche die Zuordnungsregeln, Testsituationen, Beobachtungsanweisungen usf. allenfalls *n o c h f e h l e n*. So entwirft der Psychologe Annahmengenüge (Konstruktionen, Modelle, ~~Vor~~stellungen) hinsichtlich der Ängstlich-

1) BUHR, M./KLAUS, G.: Philosophisches Wörterbuch, Berlin 1972⁸, S. 721

2) BUHR, M./KLAUS, G., a.a.O., S. 721

lichkeit und sucht sich erst im Anschluß daran sogenannte 'Operationalisierungen', mit deren Hilfe er seine Konstruktionen an der Erfahrung mißt."¹⁾

Daß dies methodologische Probleme aufwirft, soll die folgende Abbildung²⁾ andeuten.



Theoretische Ebene: Bereich des nicht unmittelbar sinnlich Wahrnehmbaren (Konstrukt 'Angst')

kein kontinuierlicher Übergang von der theoretischen zur empirischen Ebene

Empirische Ebene: Bereich des Wahrnehmbaren (Angekreuzte Antworten auf der Prüfungsskala 'TASC')

Um das Konstrukt 'Angst' wenigstens partiell zu messen, wird eine Skala konstruiert, deren items gültig für die zu messende Dimension sein sollen. In der Testtheorie bezeichnet Gültigkeit "den Grad der Genauigkeit (an), mit dem dieser Test dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise, das (die) er messen soll oder zu messen vorgibt, tatsächlich mißt."³⁾

Es handelt sich beim Gültigkeitsproblem also um ein Repräsentationsproblem.

Das methodologische Problem besteht - mathematisch formuliert - darin, daß die Zuordnung Konstrukt/ Beobachtungs- oder Meßverfahren nicht uneindeutig ist. Grund dafür und somit "Ausgangs-

1) HERRMANN, Th.: Persönlichkeitsmerkmale, Bestimmung und Verwendung in der psychologischen Wissenschaft, Stuttgart/Berlin/Köln/Mainz 1973, S. 135

2) Wir kombinieren und modifizieren die Abbildungen von BERGER, H.: Erfahrung und Gesellschaftsform. Methodologische Probleme wissenschaftlicher Beobachtung, Stuttgart/Berlin/Köln/Mainz 1973, S. 101 und HOLM, K.: Zuverlässigkeit von Skalen und Indizes, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 1970, S. 694

3) LIENERT, G.A.: Testaufbau und Testanalyse, Weinheim 1967²⁾, S. 16

problem der Validitätsmessung ist der Tatbestand, daß für theoretische Konstrukte eindeutige Operationalisierungsvorschriften fehlen, also die Konstrukt und Beobachtung verbindenden Korrespondenzregeln unbestimmt bleiben. Damit kann sich die Prüfung von H_1 über gesetzmäßige Verbindungen von C_1 mit anderen Konstrukten *n i c h t* auf gesicherte Hypothesen stützen.¹⁾ Eine Lösung auf logischer Ebene in Form einer Deduktion²⁾ ist nicht möglich. Das gibt auch der empirisch arbeitende Sozialforscher zu: "Es gibt also keine *a b s o l u t e* Lösung des Gültigkeitsproblems. Die Gültigkeit einer Skala/Index ist nicht *b e w e i s b a r*. Was man jedoch tun kann, ist, mit Hilfe bestimmter Verfahren nachzuweisen, daß es *p l a u s i b e l* ist, für eine Skala/Index Gültigkeit anzunehmen."³⁾ Plausibilität ist jedoch logisch unbefriedigend, und Verfahren können das Gültigkeitsproblem deswegen nicht lösen, weil vor der Anwendung von statistischen Verfahren die hermeneutische Identifikation der gültigen Items geleistet werden muß

Die Auswahl valider Indikatoren wird in der Regel Experten-
gruppen überlassen, so daß letzten Endes "Entscheidungsinstanz für Gültigkeitsmessungen die Urteile sozialer Gruppen sind"⁴⁾, darüber kann auch nicht die Extraktion bestimmter Faktoren mittels der Faktorenanalyse hinwegtäuschen; denn entscheidend für die Anzahl der gefundenen Faktoren ist der *laput* durch die Experten.

2.1.2. Relevanz positivistischer Forschungsdesigns vs. action research

Die Gewinnung von Daten wird in der empirischen Sozialforschung durch Verfahren reguliert, die zugleich den "Typus von Erfahrung ... selbst definieren. Einzig die kontrollierte Beobachtung

1) BERGER, H., a.a.O., S. 101

2) Die gleiche Problematik ergibt sich bei der Deduktion von Lernzielen. Dazu: MEYER, H.L.: Einführung in die Curriculum-Methodologie, München 1972

3) HOLM, K., a.a.O., S. 694

4) BERGER, H., a.a.O., S. 103

physischen Verhaltens, die in einem isolierten Feld unter reproduzierbaren Umständen von beliebig austauschbaren Subjekten veranstaltet wird, scheint intersubjektiv gültige Wahrnehmungsurteile zu gestatten."¹⁾ Diesen restringierten Erfahrungsbegriff, der außerdem die befragten Schüler zum Objekt des empirisch arbeitenden Erziehungswissenschaftlers macht, gibt die Aktionsforschung zugunsten eines kommunikativen Ansatzes auf. Ihre Postulate lassen die Alternative zur herkömmlichen Sozialforschung erkennen:

"a) Die Problemauswahl und -definition geschieht nicht vorrangig aus dem Kontext wissenschaftlicher Erkenntnisziele, sondern entsprechend konkreten gesellschaftlichen Bedürfnissen.

b) Das Forschungsziel besteht nicht ausschließlich darin, soziologische theoretische Aussagen zu überprüfen oder zu gewinnen, sondern darin, gleichzeitig praktisch verändernd in gesellschaftliche Zusammenhänge einzugreifen.

c) Die im Forschungsprozeß gewonnenen Daten werden nicht mehr als isolierte Daten 'an sich' angesehen, sondern als Momente eines prozeßhaften Ablaufes interpretiert; sie gewinnen ihren Sinn auf der theoretischen Ebene dadurch, daß sie stets mit dem realen Prozeß als Gesamtheit zusammengedacht werden, und erhalten ihre Relevanz auf der praktischen Ebene als konstitutive Momente weiterer Prozeßabläufe.

d) Die als Problem aufgenommene soziale Situation wird als Gesamtheit - als soziales Feld - angesehen, aus der nicht aufgrund forschungsimmanenter Überlegungen einzelne Variablen isoliert werden können.

e) Die praktischen und theoretischen Ansprüche des action research verlangen vom Forscher eine zumindest vorübergehende Aufgabe der grundsätzlichen Distanz zum Forschungsobjekt zugunsten einer bewußt einflußnehmenden Haltung, die von teilnehmender Beobachtung bis zur aktiven Interaktion mit den Beteiligten reicht.

1) HABERMAS, J.: Analytische Wissenschaftstheorie und Dialektik. Ein Nachtrag zur Kontroverse zwischen Popper und Adorno, in: TOPITSCH, E. (Hrsg.): Logik der Sozialwissenschaften, Köln/Berlin 1968, S. 293

f) Entsprechend soll sich auch die Rolle der Befragten und Beobachteten verändern und ihr momentanes Selbstverständnis so festgelegt werden, daß sie zu Subjekten im Gesamtprozeß werden."¹⁾

An die Stelle des Suchens nach nomologischen Hypothesen tritt die Veränderung von Praxis. Der instrumentale Charakter des nomothetischen Gesetzesbegriffs und die damit verbundene Nichtbeachtung des gesellschaftlichen Rahmens werden auch von anderer Seite kritisiert: "Es lassen sich Anhaltspunkte nur für eine immanente Anwendung der Gesetzhypothesen ableiten, in dem Sinne, daß sie Wege zu Zielen aufzeigen, die sich außerhalb der wissenschaftlichen Aussage, aber innerhalb des gesellschaftlichen Rahmens befinden. Für die Veränderung des Rahmens selbst liefern sie überhaupt keine Anhaltspunkte."²⁾

Der Hauptvorwurf gegen empirisch-analytische Methoden besteht darin, daß sie die Emanzipation der Versuchspersonen verhinderten, weil die Versuchsteilnehmer zum Objekt des Forschers gemacht würden. Dies kommt indirekt in der Kritik BENNERS an LEMPERT zum Ausdruck: "Wo Emanzipation auf eine Außeranwendungsetzung varianter Gesetzmäßigkeiten reduziert wird, verliert sie die Bestimmung einer handlungswissenschaftlichen Kategorie und verfällt den Beschränktheiten szientifischer Technologie, deren Überwindung der Rekurs auf das emanzipatorische Interesse eigentlich leisten sollte."³⁾

Daß jedoch auch der Ansatz der Aktionsforschung nicht die ultima ratio darstellt, sehen auch ihre Befürworter. Sie erkennen "die Gefahr, daß das Handeln als solches zur Legitimation für das Ausklammern der Reflexion über gesellschaftliche Zusammenhänge"⁴⁾ führen könnte. Hier weist Aktionsforschung vom marxistischen Standpunkt aus außerdem das Defizit gesellschaftstheoretischer,

-
- 1) KLUVER, J./KRUGER, H.: Aktionsforschung und soziologische Theorien. Wissenschaftstheoretische Überlegungen zum Erkenntnisinteresse in der Aktionsforschung, in: HAAG, F.u.a.: Aktionsforschung. Forschungsstrategien, Forschungsfelder und Forschungspläne, München 1972, S. 69 ff.
 - 2) DICK, F.: die widersprüche des nomothetischen gesetzesbegriffs der empirisch-analytischen sozialwissenschaften und dessen doppelt ideologische funktion ..., giessen 1972, S. 89
 - 3) BENNER, D.: Hauptströmungen der Erziehungswissenschaft. Eine Systematik traditioneller und moderner Theorien, München 1973, S. 310
 - 4) KLUVER, J./KRUGER, H. u.a., a.a.O., S. 79

Ziele bestimmender Reflexion auf. Dieser Mangel kann nur behoben werden, wenn "strukturelle gesellschaftliche Bestimmungen erfaßt (werden), die sich denkunabhängig (unbewußt und nicht-intentional) durchsetzen. Eine Sozialforschung, die sich auf die Marxsche Theorie beziehen will, kann auf der methodologischen Ebene nicht bei der Kritik der vernachlässigten subjektiven Dimension in der traditionellen Sozialforschung stehen bleiben; sie muß sich vielmehr auch auf einen kritischen Begriff des Objektiven beziehen."¹⁾

Relevant für die Begründung unseres methodischen Zugriffs, der die Versuchspersonen zum Objekt macht, wird die von der Aktionsforschung intendierte Auflösung der Subjekt-Objekt-Relation zugunsten einer Subjekt-Subjekt-Relation. Dagegen sind wir der Meinung, daß prinzipiell jeder methodische Zugang eine Subjekt-Objekt-Relation konstituiert. Wird dies ignoriert, gibt man das Kriterium der Intersubjektivität preis und bewegt sich auf dem Felde unkritisierbarer Evidenzgefühle. "Wir können vielleicht 'subjekthaft' miteinander weinen, lachen, spielen oder einen 'freien Diskurs' miteinander führen, aber schon arbeiten können wir nicht mehr, wenn die 'Subjekthaftigkeit' nicht partiell reduziert wird."²⁾

2.1.5. Induktiv-statistische Systematisierungen

Allgemeine Hypothesen, Gesetze oder theoretische Annahmen (G) sind Bestandteil des Explanans deduktiv-nomologischer Erklärungen³⁾. Sie haben folgende Struktur:

$G_1, G_2, \dots, G_n$ (nomologische Hypothesen)	Explanans
$A_1, A_2, \dots, A_n$ (Antezedenzbedingungen)	
E	(singulärer Satz)
	Explanandum

1) MÜLLER, U.: Zur möglichen Bedeutung der empirischen Sozialforschung für die Marxsche Theorie, in: BACKHAUS, H.-G. u.a. (Hrsg.): Gesellschaft - Beiträge zur Marxschen Theorie 5, Frankfurt a.M. 1975, S. 228

2) MOLLENHAUER, K./RITTELMEYER, Ch.: "Empirisch-analytische Wissenschaft" versus "Pädagogische Handlungsforschung": eine irreführende Alternative, in: Zeitschrift für Pädagogik 1975, S. 690

3) STEGMÜLLER, W., a.a.O., S. 86

Die formallogische Schreibweise würde deutlicher die deterministische Natur von DN-Erklärungen zeigen. In der empirischen Unterrichtsforschung hat man es jedoch mit statistischen Gesetzmäßigkeiten zu tun. Das zu erklärende Ereignis tritt nur mit einer gewissen Wahrscheinlichkeit ein. Deterministische Gesetzmäßigkeiten können dabei als Sonderfall von probabilistischen interpretiert werden: $p=1$ oder $p=0$.

Vom statistischen ist der mathematische Wahrscheinlichkeitsbegriff abzugrenzen, der "nur eine partielle Explikation des statistischen ... (leistet), der für die Anwendung nicht ausreicht. (Er) ... ist eine mathematische Entität. Die mit seiner Hilfe formulierten Wahrscheinlichkeitsaussagen sind rein mathematische Sätze. Der in statistischen Hypothesen verwendete Wahrscheinlichkeitsbegriff dagegen ist empirisch. (Er) ... ist eine empirische Entität. Die mit seiner Hilfe formulierten Wahrscheinlichkeitsaussagen sind rein empirische Sätze." ¹⁾ Wahrscheinlichkeit eines Ereignisses bedeutet mathematisch die relative Häufigkeit seines Vorkommens, bezogen aber auf eine unendliche Zahl von Fällen.

Ebenso unbrauchbar wie der mathematische Wahrscheinlichkeitsbegriff ist auch der statistische Syllogismus für die empirische Forschung; denn man kann zeigen, daß aus wahren Prämissen gefolgerte Konklusionen sich paralysieren. ²⁾

Die Hypothesenbildung in der empirischen Forschung geschieht nun nicht in zufälliger Weise, sondern sie wird durch die Auswahl relevanter Variablen determiniert, so daß "mit statistischen Hypothesen bei vorliegenden Randbedingungen das Auftreten irgendwelcher Ereignisse ausgeschlossen" ³⁾ wird. Dieser Tatbestand wird von induktiv-statistischen Systematisierungen berücksichtigt. "Die logischen Widersprüche lassen sich beheben, wenn man eine Reinterpretationen der Wendungen 'es ist sehr wahrscheinlich,

1) STEGMÜLLER, W., a.a.O., S. 642

2) STEGMÜLLER, W., a.a.O., S. 634

3) OPP, K.-D.: Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theoriebildung, Hamburg 1970, S. 39

daß ...', 'es ist praktisch sicher, daß ...' oder 'die Wahrscheinlichkeit dafür, daß ..., ist ...' in den Konklusionen (Explananda) vornimmt."¹⁾ Diese Wendungen qualifizieren dann die Konklusio nicht mehr als solche, sondern sie beschreiben die Relation, die zwischen den Prämissen und der Konklusio besteht.²⁾

Für induktiv-statistische Systematisierungen gelten folgende Adäquatheitsbedingungen:

- "B₁₊. Das Argument, das vom Explanans zum Explanandum führt, muß induktiv korrekt sein. Die induktive Wahrscheinlichkeit des Explanandums relativ zum Explanans muß im Rahmen der induktiven Logik bzw. der normativen Theorie der personellen Wahrscheinlichkeit ... korrekt berechnet sein.
- B₂₊. Das Explanans muß mindestens ein allgemeines statistisches Gesetz enthalten (oder einen Satz, aus dem eine allgemeine Wahrscheinlichkeitshypothese ableitbar ist).
- B₃. Das Explanans muß empirischen Gehalt besitzen.
- B₄₊. Die Sätze, aus denen das Explanans besteht, müssen wahr sein.
- B₄. Die Sätze, aus denen das Explanans besteht, müssen gut bewährt bzw. bestätigt sein."³⁾

2.2. Probleme der Methodik

Den Gegenstand unserer Untersuchung versuchen wir mit Hilfe bestimmter Methoden in den Griff zu bekommen. Zu dem verwendeten Instrumentarium gehören die Einstellungsskala 'Englisch', das Semantische Differential, der Fremdspracheneignungstest, das Leistungsprüfsystem (LPS), Angstskalen und die informellen Tests I und II.

Da bei der Beschreibung dieser Methoden auch ihre speziellen methodologischen Probleme diskutiert werden, genügt es an dieser Stelle, die methodenimmanenten Probleme stichwortartig an-

1) WESTMEYER, H.: Kritik der psychologischen Unvernunft. Probleme der Psychologie als Wissenschaft, Stuttgart/Berlin/Köln/Mainz 1973, S. 22

2) STEGMÜLLER, W., a.a.O., S. 651

3) WESTMEYER, H., a.a.O., S. 26 - 27

zudeuten. Ist die Einstellungsskala 'Englisch' eindimensional? Welche Rolle spielt bei ihr und den Fragebogen der Bedeutungsaspekt? Ist Äquidistanz der Intervalle beim Semantischen Differential gewährleistet? Welcher Bedeutungsverlust geht zu Lasten der Operationalisierungen? Welche Implikationen ergeben sich aus dem Verknüpfungssaxiom der klassischen Testtheorie?

3. Problemstellung

Die Fragestellung unserer Untersuchung richtet sich nicht auf den Nachweis, ob z.B. informelle Tests für die Leistungsüberprüfung effektiver sind als herkömmliche Klassenarbeiten, sondern auf die Nebenwirkungen dieser beiden Prüfinstrumente. Die Meßergebnisse der abhängigen Variablen sind deshalb auch nicht die Produkte kognitiver Leistungen, sondern die quantifizierten affektiven Eindrücke, welche die Versuchsgruppen auf die Konzepte 'Selbstbild', 'Wunschbild', 'schwacher Schüler', 'guter Schüler', 'Englischlehrer', 'Klassenarbeit', 'Zensur' und 'Klasstest' projiziert haben. Damit lenkt unser Forschungsinteresse die Aufmerksamkeit auf die emotionalen Komponenten von Einstellungen, auf den affektiven Einfluß graduell verschieden strukturierter Prüfinstrumente normorientierter Leistungsmessung.

Dabei vermuten wir, daß der Strukturierungsgrad (Gestaltprägnanz) normorientierter Prüfinstrumente zu unterschiedlichen Effekten bei der Einstufung der Konzepte durch wenig ängstliche oder sehr ängstliche Schüler führt. Begründet werden kann diese Generalhypothese mit dem aus der Wahrnehmungspsychologie bekannten Figur-Grund-Prinzip. Da aber die Fähigkeit, differenziert wahrzunehmen, eine Funktion des Alters ist, wurden in das experimentelle Design der Pilotstudie zwei Jahrgangsstufen einbezogen.

Unser weiteres Interesse richtet sich darauf festzustellen, welche wertgetönten Bereitschaften, Dispositionen - notwendige, jedoch nicht hinreichende Bedingungen für Verhaltensweisen - durch normorientierte Prüfinstrumente ausgelöst werden. Die auf der Bewertungsdimension des Semantischen Differentials gemessenen Einstellungen als Bedingungen der Möglichkeit von Verhalten werden durch die polaren Adjektivpaare operational in zwei Klassen geteilt. Die Urteile auf der negativen Skalseite bezeichnen wir als 'dissoziativ', die auf der positiven als 'integrativ'. Diese Operationalisierung wurde auch auf den

Dimensionen 'Stärke' und 'Aktivität' vorgenommen.

Die kognitive Dissonanz zwischen der Propagierung informeller Tests - auch ihrer unkritischen Übernahme durch Gesamtschulen - und dem Fehlen empirischer Untersuchungen über kurz- oder langfristige Wirkungen der verwendeten Methoden der Leistungsüberprüfung auf die Schülerpsyche löste die Problemstellung aus; zumal die Analyse der Syntax informeller Tests eine Reihe hypothetisch fragwürdiger Momente ihres Kalküls ergab, die den vorteil der Leistungsobjektivierung einzuschränken schien. Doch abgesehen von der subjektiven Problemgenese macht der objektive Stand der Prüfungsforschung es erforderlich, die auf diesem Felde geäußerten Vermutungen mit einem empirischen Kriterium zu konfrontieren. "Die Auswirkungen von Prüfungen (Tests) wurden in Deutschland noch nicht systematisch untersucht. Berichte darüber stellen hauptsächlich Spekulationen oder Verallgemeinerungen von Einzelbeobachtungen dar."¹⁾

1) KRUMM, V.: Artikel 'Leistung - Schulleistung - Schulleistungsmessung', in: WULF, Chr. (Hrsg.): Wörterbuch der Erziehung, München/Basel 1974, S.386

4. FORSCHUNGSDESIGN, -METHODE UND STATISTISCHES INSTRUMENTARIUM DER UNTERSUCHUNG. METHODISCHE UND METHODOLOGISCHE PROBLEME

Die allgemein gehaltene wissenschaftstheoretische Auseinandersetzung mit einigen Theorie-Begriffen und deren Implikationen in der deutschen Soziologie¹⁾ wurde bald von anderen Fachdisziplinen²⁾ aufgenommen. Der Rezeption in der Erziehungswissenschaft folgte die Problematisierung empirisch-analytischer Methoden. Hatte z.B. BLANKERTZ³⁾ noch nach dem grundsätzlichen Zusammenhang von pädagogischer Theorie und empirischer Forschung gefragt, so zeichnete sich später eine speziellere Diskussion über Möglichkeiten und Grenzen empirisch-analytischer Verfahren ab. ACHTENHAGEN kritisierte Fachdidaktiken unter ideologiekritischem, wissenschaftstheoretischem Aspekt, da Entscheidungen im Unterricht und bei der Lehrplangestaltung durch Normen affiziert werden. Deshalb müsse das dahinter stehende Wertsystem reflektiert werden. Letzteres kann der empirisch-analytische Ansatz

-
- 1) ADORNO, Th. W. (Hrsg.): Der Positivismusstreit in der deutschen Soziologie, Neuwied und Berlin 1972²
FIJALKOWSKI, J.: Über einige Theorie-Begriffe in der deutschen Soziologie der Gegenwart, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 1961, S. 88 - 109
 - 2) BRENTANO, M. von: Die unbescheidene Philosophie. Der Streit um die Theorie der Sozialwissenschaften, in: Argument 43, 1967, S. 102 - 116
HOLZKAMP, K.: Wissenschaftstheoretische Voraussetzungen kritisch-emanzipatorischer Psychologie, in: Zeitschrift für Sozialpsychologie, 1970, S. 5 - 21 und S. 109 - 141
ACHTENHAGEN, F.: Erziehungswissenschaft und empirisch-analytische Wissenschaftstheorie, in: Pädagogische Rundschau, 1969, S. 840-856
SEILER, H.: Moderne Forschungstheorie und Erziehungswissenschaft. Bemerkungen zum Programm der empirischen Pädagogik, in: Zeitschrift für Pädagogik, 1969, S. 501 - 528
WELLENDOFF, F.: Ansätze zur erziehungswissenschaftlichen Theoriebildung in der Bundesrepublik, in: FURCK, C.-L./GOLDSCHMIDT, D./RÖBBELEN, I. (Hrsg.): Gesellschaft und Erziehung I, Erziehungswissenschaft als Gesellschaftswissenschaft, Heidelberg 1969, S. 68 - 110
MEILE, B.: Voraussetzungen der empirischen Pädagogik. Ein wissenschaftstheoretischer Aufriß, in: Zeitschrift für Pädagogik, 1972, S. 709 - 733
 - 3) BLANKERTZ, H.: Pädagogische Theorie und empirische Forschung, in: Neue Folge der Ergänzungshefte zur Vierteljahresschrift für wissenschaftliche Pädagogik, Heft 5, 1966, S. 65 - 78

nicht leisten: "Der rein empirisch-analytische Ansatz ist damit aufgehoben in der dialektischen Ideologiekritik."¹⁾ So hat MEYER²⁾ herausgearbeitet, was erfahrungswissenschaftliche Techniken für curriculare Entscheidungen zu leisten vermögen. THIEMANN analysierte empirisch-pädagogische Untersuchungen auf ihre Relevanz "zu einer Erweiterung erfolgreich planenden Handelns" und "zu einer Erneuerung von Unterricht"³⁾.

Diesen Diskussionsstand kann der empirisch arbeitende Erziehungswissenschaftler nicht ignorieren, wenn er sich fragt, welches Forschungsinteresse ihn bei seiner Untersuchung leitet, welche Bedeutsamkeit arithmetische Mittel und Standardabweichungen für einen individualisierten Unterricht besitzen, ob der Schüler als Subjekt oder Objekt erzieherischen Handelns gesehen wird, inwieweit die Zielvorstellung 'Mündigkeit' eine Beteiligung des Edukanden am Forschungsprozeß verlangt.

Approximativ versuchen wir, den Forderungen SEILERS nachzukommen, der zur Bedingung empirischer Forschungstätigkeit macht, daß "sich einerseits jeder inhaltliche Beitrag zur empirischen Pädagogik mindestens in vierfacher Weise zu rechtfertigen (hat):

- a) er muß dem erreichten Stand der empiristischen Wissenschaftstheorie entsprechen;
- b) er muß mit dem z.Z. verfügbaren und akzeptierten Instrumentarium der erfahrungswissenschaftlichen Forschungspraxis prüfbar sein;
- c) er muß aufweisbar der augenblicklichen Erziehungspraxis als seinem "Gegenstand" adäquat sein und zur Lösung praktisch-pädagogischer Probleme beitragen;
- d) er muß schließlich seinen Ort in der umgreifenden Sinndeutung und kritik angeben können, die in der traditionellen Reflexion auf die pädagogische Praxis

1) ACHTENHAGEN, F.: Didaktik des fremdsprachlichen Unterrichts, Weinheim 1969, S. 41

2) MEYER, H.L.: Curriculum - Methodologie, München 1969

3) THIEMANN, F.: Der Beitrag empirischer Unterrichtsforschung für die Konzeption von Unterricht, Bad Heilbrunn 1973

und deren gesellschaftlich-geschichtliche Bezüge vorliegen."¹⁾

Entscheidungshilfe bei der Passung Methode/Gegenstand unserer Untersuchung gibt die wissenschaftstheoretische Auseinandersetzung mit den gängigen Methoden der empirischen Sozialforschung. Geht man davon aus, daß das "p ä d a g o g i s c h e F e l d (so) z u s t r u k t u r i e r e n (sei), d a m i t d i e V e r n ü n f t i g k e i t d e r z u e r z i e h e n d e n S u b j e k t e n i c h t v e r h i n d e r t, s o n d e r n g e f ö r d e r t w e r d e"²⁾, dann erweisen sich die Forschungstechniken empirischer Sozialforschung, die aufgrund ihres Rationale das Individuum ausblenden, indem sie es mit Durchschnittswerten beschreiben, als defizitär; denn standardisierte untersuchungsmethoden abstrahieren von der Situation und gesellschaftlichen Lebensform. Konstitutiv für den methodologischen Rahmen von action research dagegen ist das Beteiligtsein der Subjekte am Forschungsprozeß, um die Subjekt-Objekt-Relation aufzuheben. Diese Forderung übersieht aber, daß "jeglicher Erkenntnisvorgang notwendigerweise das Zum-Objekt-Machen des Erkenntnisgegenstandes impliziert ... Wie weit sie in einem Forschungsprozeß reduziert werden s o l l t e, dafür kann es keine allgemein verbindliche Anweisung geben; es hängt ab vom Forschungsziel und der Beschaffenheit des 'Gegenstandes'."³⁾

Ziel unserer Pilostudie ist es, die affektiven Wirkungen einer restringierten, nur asymmetrische Kommunikation zulassenden normorientierten Prüfungssituation festzustellen, wie sie sich in der Beurteilung relevanter Konzepte des schulischen sozialen Feldes sedimentieren. Da nun die für die Prüfungssituation konstitutive Aufgabenstellung der herkömmlichen schulischen Prüfungsmethodik den Schüler prinzipiell zum Objekt macht, erscheint die Anwendung empirisch-analytischer Methoden legitim. Der Zu-

1) SEILER, H.: Moderne Forschungstheorie und Erziehungswissenschaft. Bemerkungen zum Programm der empirischen Pädagogik, Zeitschrift für Pädagogik, 1969, S. 519

2) MOLLENHAUER, K.: Pädagogik und Emanzipation. Polemische Skizzen, München 1968, S. 70

3) MOLLENHAUER, K./RITTELMAYER, Ch.: "Empirisch-analytische Wissenschaft" versus "Pädagogische Handlungsforschung": eine irreführende Alternative, in: Zeitschrift für Pädagogik, 1975, S. 690

griff durch action research wäre adäquater, wenn die Subjekte durch Gegenfragen die Aufgabenstellung des Prüfers verändern könnten. Die aufgrund unserer Methodik gefundenen empirischen Beziehungen zwischen Variablen dürfen jedoch nicht hypostasiert werden, da eine pädagogische Empirie "die Erziehungswirklichkeit nicht mehr in den Beziehungen der auf Bedingungsfaktoren und Variablen von Gesetzeszusammenhängen reduzierten Gegebenheiten" auffaßt, sondern "die Erziehungswirklichkeit vielmehr als Ort der Verwirklichung sinnbestimmter Erziehungspraxis" begreift und "sich als praxeologisch-erfahrungswissenschaftliche Planung und Kontrolle von Versuchen einer an den Möglichkeiten und Aufgaben der Erziehung ausgerichteten Erziehungspraxis"¹⁾ zu verstehen sucht.

4.1. Zum Begriff des pädagogischen Experiments

Hauptsächlich zwei Problemebenen werden bei den folgenden Ausführungen anvisiert:

- a) ein methodenimmanenter Problembereich, der z.B. die formale Struktur des naturwissenschaftlichen Experiments, Definitionen des Begriffs 'Experiment' usw. betrifft und
- b) ein methodologischer Problembereich, der z.B. die Problematisierung des naturwissenschaftlichen Gesetzesbegriffs, der die Behandlung des Individuums als Objekt im technologischen Experiment zum Gegenstand hat.

Die Logik des naturwissenschaftlichen Experiments unterscheidet sich von der des sozialwissenschaftlichen aufgrund der unterschiedlichen Struktur der beiden Objektbereiche.²⁾ Genügt beim physikalischen Experiment eine einfache Vorher - Nachheranordnung, so ist bei der sozialwissenschaftlichen Versuchsanordnung eine Parallel- oder Kontrollgruppe die *conditio sine qua non*, da es schwierig ist, die Versuchsbedingungen konstant zu halten. Nach CAMPBELL sind es acht Faktoren, welche die *innere Gültigkeit* des Experiments gefährden können: Zwischen-

1) BENNER, D.: Hauptströmungen der Erziehungswissenschaft. Eine Systematik traditioneller und moderner Theorien, München, 1973, S. 158

2) Je nach der wissenschaftstheoretischen Position wird man dieser Behauptung zustimmen oder sie ablehnen

Die logische Stringenz des Beweisganges sollte den Erziehungswissenschaftler aber nicht daran hindern zu fragen, welche Implikationen die Bevorzugung dieser experimentellen Anordnung für eine Unterrichtsforschung hat, die sich der präskriptiven Norm 'Mündigkeit' verpflichtet weiß und sie in das Arrangement von Untersuchungsmethoden eingehen lassen möchte.¹⁾

Ebenso kritisch sind die Adäquatheitsbedingungen für Kontrollgruppenexperimente zu untersuchen:

- "a) Die Ausgangssituationen als Experimentalgruppe und Kontrollgruppe, müssen gleich sein, und zwar hinsichtlich aller Faktoren.
- b) Die Ausgangssituationen müssen gleichzeitig gebildet werden, und auch die Nachhermessung von Experimental- und Kontrollgruppe nach Einwirkung des experimentellen Faktors muß gleichzeitig geschehen.
- c) Während des Experiments muß die Kontrollgruppe gegen einen Einfluß des experimentellen Faktors vollkommen abgeschirmt sein.
- d) Alle äußeren und inneren Bedingungen des Experiments müssen für Experimental- und Kontrollgruppe während des ganzen Versuchs völlig gleich sein; beide Gruppen dürfen sich nur durch den experimentellen Faktor unterscheiden."²⁾

Die Struktur des Kontrollgruppenexperiments scheint die Selektion bestimmter Forschungsprobleme zu protegieren, indem die Wahrnehmung mehr auf die Lern p r o d u k t e als auf die Lern p r o z e s s e ³⁾ gerichtet wird. Plausibel erscheint auch die Strukturierung des Experimentierfeldes durch unabhängige und abhängige Variablen, analog dem physikalischen Experimentieren. Problematisch wird die Übernahme dieses naturwissenschaftlichen Paradigmas für die sozialwissenschaftliche Forschung, wenn man die "positivistische Gegenthese der logischen Einheit von theoretischen und historischen Wissenschaften"⁴⁾ nicht akzeptiert. Die Gefahr liegt nahe, daß man sich der 'black-box-Strategie' orthodoxer Behavioristen anschließt, Reaktionen aus einem stimuli-set

-
- 1) THIEMANN, F.: Der Beitrag empirischer Unterrichtsforschung für die Konzeption von Unterricht, Bad Heilbrunn 1973, S. 150
 - 2) SCHULZ, W., a.a.O., S. 101
 - 3) Ein Vorwurf, den THIEMANN der empirischen Unterrichtsforschung macht, a.a.O., S. 152
 - 4) HABERMAS, J.: Zur Logik der Sozialwissenschaften, Frankfurt am Main 1970, S. 125

zu erklären, dabei jedoch intentionales Handeln auf reflexhaftes Verhalten reduziert. "Der behaviouristische Ansatz schränkt die Wahl der theoretischen Grundannahmen so ein, daß sich Gesetzhypothesen auf einen Zusammenhang von Stimuli und Verhaltensreaktionen beziehen lassen, während der aktionistische Ansatz einen kategorialen Rahmen festlegt, innerhalb dessen Aussagen über intentionales Handeln gemacht werden."¹⁾

HOLZKAMP zielt mit seiner Kritik auf die im Experiment kontrollierten und standardisierten Ausgangsbedingungen, indem er den gesellschaftlichen Rahmen thematisiert. Er behauptet, daß "die in der Psychologie vom Experimentator hergestellten Ausgangsbedingungen, anders als in der Physik, kaum geeignet sind, als einheitsstiftendes Moment für die psychologische Theorie zu dienen, weil sie gar nicht die eigentlichen Bedingungen für das Zustandekommen der Reaktionen der Vpn als experimenteller Effekte sind. Diese Bedingungen liegen vielmehr in verborgenen Sichtweisen und Stellungnahmen der Vpn selbst."²⁾ WESTMEYER hält diese Behauptung für unverbindlich und meint, "ob in jedem Fall das Verhalten einer Person in der experimentellen Situation durch ihre konkrete historisch-gesellschaftliche Lage, in der sie sich außerhalb des Laboratoriums befindet, bedingt wird oder ob die experimentelle Anordnung die wesentlichen Determinanten enthält, ist eine empirische Frage."³⁾

Festzuhalten ist jedoch, daß von den Vpn eine Interpretationsleistung bei der Identifizierung des 'sozialen Objekts' verlangt wird, dessen operationale Gestalt z.B. in Form eines Fragebogens oder einer Einstellungsskala vorliegt; aber: "Die Beziehung zwischen Zeichen und sozialem Objekt ist keine eindeutige. Verbale Instruktionen mögen standardisiert erscheinen ... aber ihr 'augenfälliger' Charakter und ihre Bedeutung können nicht als selbstverständlich gegeben erachtet werden."⁴⁾ Der Wahrnehmungs-

1) HABERMAS, J., a.a.O., S. 139

2) HOLZKAMP, K.: Zum Problem der Relevanz psychologischer Forschung für die Praxis, in: Psychologische Rundschau, 1970, S. 6

3) WESTMEYER, H.: Kritik der psychologischen Unvernunft. Probleme der Psychologie als Wissenschaft, Stuttgart/Berlin/Köln/Mainz 1973, S. 127

4) CICOUREL, A. von: Methode und Messung in der Soziologie, Frankfurt am Main 1970, S. 225

prozeß scheint daher mit großer Wahrscheinlichkeit durch die Klassenlage determiniert. "Die Ermittlung des praktischen Bedeutungskreises sprachlicher Ausdrücke führt also zu grundlegend anderen Ergebnissen als eine inhaltliche Bedeutungsanalyse."¹⁾

Die Messung der Ausprägungsgrade der abhängigen Variablen mit Hilfe von Prüfverfahren bringt zwar nicht ein grundsätzliches, doch ein forschungspraktisch nicht zu unterschätzendes Problem auch dadurch mit sich, "daß sie (die Prüfverfahren - K.H.) von den Versuchspersonen K o o p e r a t i o n verlangen. Der Sozialforscher, der einen Probanden testen oder befragen möchte, ist in mehrfacher Hinsicht auf Kooperation angewiesen:

- a) er muß das Einverständnis der Versuchspersonen haben, seine Fragen stellen zu dürfen,
- b) er muß erwarten können, daß die Versuchsperson seiner Frage zuhört,
- c) er muß auf Bereitwilligkeit der Versuchsperson hoffen, auf die Frage zu antworten und
- d) er sollte Gewißheit haben, daß die Antwort auch ernst und ehrlich gemeint ist."²⁾

Trotz uneinheitlicher Verwendung des Terminus 'Experiment' in der psychologischen Literatur³⁾ wird doch in den meisten Definitionsversuchen das Testen einer Kausalhypothese als wesentliches Merkmal des experimentellen Vorgehens herausgestellt. "Die essentiellen Kriterien für jedes Experiment sind die Kausalhypothese, vergleichbare Situationen und Kontrolle der Faktoren, wobei die Situation definiert sein muß durch eine E x p e r i m e n t i e r g r u p p e, die einer bestimmten Behandlung unterworfen wird, und durch eine K o n t r o l l g r u p p e, die keine solche Behandlung erfährt. Bei beiden Gruppen werden Beobachtungen oder Messungen mit einem Kategoriensystem, Graduierungsskalen oder

1) BERGER, H., a.a.O., S. 122

2) SCHULZ, W., a.a.O., S. 74 - 75

3) vgl. BREDEKAMP, J.: Experiment und Feldexperiment, in: GRAUMANN, C.-F.(Hrsg.): Sozialpsychologie, Hdb. d. Ps., Bd. 7 (1. Halbband), Göttingen 1969, S. 332

anderen Meßinstrumenten während des Ablaufs des Experiments vorgenommen."¹⁾ Wie aus der schematischen Darstellung des Kontrollgruppenexperiments zu ersehen ist, ist für dieses der Zeitfaktor t konstituierend. "Da Kausalprozesse bzw. Wechselwirkungen in der Zeit verlaufen, sind solche Zusammenhänge strenggenommen nur im Experiment überprüfbar, in welchem die Zeit als variable explizit im Modell der Anordnung aufgenommen ist. Diese Struktur weisen auch alle Längsschnittuntersuchungen auf. Alle Querschnittuntersuchungen - und die quasiexperimentellen Anordnungen mit Versuchs- und Kontrollgruppe - liefern *keine* Kausalbeziehungen, sondern nur Zusammenhänge in der Form von Kovariationen bzw. Korrelationen von Merkmalsdimensionen. Die kausale Interpretation von solchen Zusammenhängen, und diese engere Bedeutung hat der Ausdruck 'Interpretation von Daten', muß von der zugrundegelegten *Theorie* geleistet werden. Das Experiment ist eine Möglichkeit, die Richtigkeit solcher Interpretationen zu prüfen."²⁾ KLAUER meint, wenn er sich zunächst auch gegen "die Einschränkung des Experiments auf die Erforschung von kausalzusammenhängen"³⁾ wendet: "Die Behauptung von Kausalzusammenhängen ist grundsätzlich eine Angelegenheit der Theoriebildung."⁴⁾

Experimentelles Vorgehen scheint demnach auf eine hoch ausgebildete Theorie zu verweisen. Die Bemerkungen verschiedener Autoren zu diesem Fragenkomplex weichen der Konsequenz dieser Forderung aus, nur dann Experimente zu veranstalten, wenn eine Theorie das feste Fundament abgibt. Man zieht sich aus der Affäre, indem man dem Experiment verschiedene Funktionen zuweist. TRAVERS unterscheidet zwischen dem Experiment als *Demonstration*, wie der Begriff früher verwendet wurde, "wobei es sich auf Erkundungen bezog, von denen man nicht behaupten kann, daß sie auf rigorose Prüfung von Hypothesen gerichtet waren" und

1) MÜNCH, W.: Datensammlung in den Sozialwissenschaften, Stuttgart/Berlin/Köln/Mainz 1971², S. 45

2) MEILE, B., a.a.O., S. 725

3)+4) KLAUER, K.: Das Experiment in der pädagogischen Forschung. Eine Einführung, Düsseldorf 1973, S. 27

Experiment als Situation, "in der eine Bedingung oder mehrere Bedingungen planvoll variiert werden, um den Effekt der Variationen untersuchen zu können."¹⁾ Nach METZGER erfüllt das Experiment ebenfalls zwei Funktionen - erkunden oder entscheiden. Außerdem betont er die Wichtigkeit der 'hermeneutischen Phase' vor allem Experimentieren.²⁾

Der Erforschung von Kausalzusammenhängen sind im schulischen Lernfeld Grenzen gesetzt, wenn man einerseits die 'natürlichen'³⁾ Schulbedingungen beibehalten will, aber andererseits die Absicht hat zu generalisieren. Höchst problematisch wirkt sich die Beibehaltung der sogenannten 'natürlichen' Situation bei innovativen Bemühungen aus, die ja ein Experiment im weiten Sinn darstellen.

ECKEL verlangt "in Übereinstimmung mit Lay .. vom pädagogischen Lernexperiment folgende Eigenschaften:

- (I) Es soll in der Schulklasse unter üblichen schulischen Bedingungen stattfinden.
- (II) Es soll sich auf einen 'sinnvollen', didaktisch zubereiteten, in der Schule üblicherweise zu unterrichtenden Lehrstoff beziehen.
- (III) Abweichungen von den Forderungen (I) und (II) sollen möglichst gering sein."⁴⁾

Der kausal-analytische Zugriff des Experiments intendiert die Aufdeckung von Gesetzmäßigkeiten in der 'Wirklichkeit'. Nomologische Hypothesen haben die logische Form: $\forall x \{Fx \rightarrow Gx\}$. "Unter einem Gesetz soll eine Aussage verstanden werden, die 1. ohne raumzeitlichen Bezug ist, in der 2. (mindestens) einer unendlichen Menge von bestimmten Objekten (mindestens) ein Merk-

1) TRAVERS, R.M.W.: Einführung in die erziehungswissenschaftliche Forschung, München 1972, S. 269 ff.

2) METZGER, W.: Das Experiment in der Psychologie, in: Studium Generale, 1952, S. 147

vgl. MAYNTZ, R. u.a.: Einführung in die Methoden der empirischen Soziologie, Köln und Opladen 1969, S. 168

3) Der Begriff erscheint uns zu vage

4) ECKEL, K.: Die Bedeutung des Klasseneffekts für die schulpädagogische Forschung, in: Programmierteres Lernen und programmierter Unterricht, 1969, S. 98 und ECKEL, K.: Das pädagogische Methodenexperiment, in: Programmierteres Lernen und programmierter Unterricht, 1970, S. 158

mal zugeschrieben wird und die 3. als empirisch wahr akzeptiert wird."¹⁾ Wissenschaftstheoretische Kritik setzt beim Allquantor ein, weil "Aussagen über empirisch vorfindbare Regelmäßigkeiten der historisch-sozialen Realität nicht durch Eliminierung aller Raum-Zeit-Begriffe strukturell relativiert und damit vorschnell 'nomologisiert' werden dürfen, wie es die empirisch-analytisch orientierte Denkweise nahelegt, sondern daß es selbst streng allgemeingültig erscheinende empirische Befunde so lange wie möglich auf ihre raum-zeitliche Bedingtheit hin befragt und historisch relativiert werden müssen."²⁾ Radikalere Kritik übt BENNER von seinem Ansatz einer gesellschaftskritischen praxeologischen Empirie. Am Beispiel der kompensatorischen Erziehung zeigt er, "daß kompensatorische Erziehung in Ermangelung einer kritischen Erziehungs- und Bildungstheorie letztlich darin aufgeht, variante Gesetzmäßigkeiten außer kraft zu setzen und gleichzeitig variante Gesetzmäßigkeiten zu produzieren. Der Tatbestand gesellschaftlicher Ungerechtigkeit wird hierdurch zu einem Tatbestand der Benachteiligung durch variante Gesetzmäßigkeiten verniedlicht. Der empirische Sozialwissenschaftler gerät hierdurch, sofern er um die technologische Auslegung seiner szientifischen Einsichten weiß, in die gleiche Situation wie der Naturwissenschaftler: um den Verwertungshorizont seiner Forschungsergebnisse einer praxeologischen Kritik zu unterziehen, muß er eingestehen, daß das von ihm gelieferte nomologische Wissen nicht zum Begriff der diesem Wissen selber vorausgesetzten Wirklichkeit hypostasiert werden darf."³⁾ Dem technologischen Experiment setzt BENNER das

-
- 1) OPP, K.-D.: Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theoriebildung, Hamburg 1970, S. 48
 - 2) KRECKEL, R.: Soziologische Erkenntnis und Geschichte. Über Möglichkeiten und Grenzen einer empirisch-analytischen Orientierung in den Humanwissenschaften, Opladen 1972, S. 119
 - 3) BENNER, D.: Hauptströmungen der Erziehungswissenschaft. Eine Systematik traditioneller und moderner Theorien, München 1973, S. 342

praktische Experiment als Alternative gegenüber. "Während das Bedürfnis nach technologischen Experimenten aus dem Widerspruch zwischen partikularen Interessen und der Wirklichkeit erwächst, z.B. aus dem Widerspruch zwischen Abnehmerbedürfnissen und Schulleistungen, hat das praktische Experiment seinen Ursprung in Widersprüchen zwischen der Idee und der Wirklichkeit, welche zunächst am Widerstreit partikularer Interessen als Widersprüche innerhalb der vorgegebenen gesellschaftlichen Situation aufscheinen. Indem im praktischen Experiment jeder technologisch reparierbare Widerspruch transzendiert und auf seinen Ursprung im Widerstreit zwischen Idee und Wirklichkeit hin untersucht wird, treten in das praktische Experiment die gesellschaftliche Wirklichkeit und die hinter dieser stehenden partikularen Interessen selber als Variable ein. Die partikularen Interessen und die von diesen beherrschte gesellschaftliche Wirklichkeit werden ihrerseits fraglich; ihr Fraglichwerden unter der regulativen Idee einer in der Erfahrung nicht vorgegebenen Vollkommenheit wird zum Inhalt des praktischen Experiments, welches an der Höherbildung der Gesellschaft und der Verwirklichung der Idee der Vollkommenheit ausgerichtet ist."¹⁾

4.2. Untersuchungshypothesen

Vor der inhaltlichen Hypothesenbeschreibung skizzieren wir die Form von Hypothesen. Sinnvoll erscheint dies deshalb, weil dadurch ein Bezugssystem entsteht, das drei Funktionen erfüllen kann:

1. Die Reichweite von Hypothesen kann besser abgesteckt werden, so daß eine Über- oder Unterinterpretation empirischer Befunde genauer abgeschätzt werden kann.
2. Es wird eine differenziertere Diskussion z.B. von Thesen der Art möglich, "daß die empirische Einzelforschung in der Tat keinen Beitrag zu einer Individualisierung von Unterricht liefert, da sie, interessiert an t r a n s s i t u a t i o n a l e n A l l a u s s a g e n (von mir hervorgehoben - K.H.), gerade die individuellen Aus-

1) BENNER, D., a.a.O., S. 344

prägungen von Verhaltensformen und Verarbeitungsweisen in ihren Entwürfen außer Ansatz läßt. Sie werden vielmehr aus dem Forschungskontext gänzlich ausgeblendet oder aber auf einen Durchschnittswert gebracht, der dann zum Bezugspunkt ihrer Konstruktionen gemacht wird."¹⁾

3. Die Anwendung statistischer Signifikanztests wird problematisiert.²⁾

Hypothesen "können als Allsatz, Existenzsatz oder Wahrscheinlichkeitshypothese formuliert werden."³⁾ Klassifiziert man nach den Merkmalen verifizierbar/falsifizierbar und den Klassen endlicher/unendlicher Individuenbereich, so ergibt sich folgender Zusammenhang:

Abb. 4: Formale Klassifizierung von Hypothesen

		verifizierbar	falsifizierbar
Allsatz	endlicher Individuenbereich	x	x
	unendlicher Individuenbereich		x
Existenz-	endlicher Individuenbereich	x	x
	unendlicher Individuenbereich	x	

BREDENKAMP unterscheidet zwischen (1) allgemeinpsychologischen Hypothesen, formuliert als unbegrenzte Allsätze, (2) differentialpsychologischen Hypothesen, formuliert als unbegrenzte Allsätze, die einen Spielraum für interindividuelle Unterschiede zulassen, und (3) statistischen Hypothesen ohne Allgemeinheitsgrad. Er veranschaulicht die drei Formen psychologischer Hypothesen am Beispiel der Wachstumsfunktion, die einen funktionalen Zusammenhang zwischen Lebensalter (X) und Intelligenzquotient (Y) behauptet,

1) THIEMANN, F.: Der Beitrag empirischer Unterrichtsforschung für die Konzeption von Unterricht, Bad Heilbrunn 1973, S. 147

2) BREDENKAMP, J.: Der Signifikanztest in der psychologischen Forschung, Frankfurt am Main 1972

3) BREDENKAMP, J., a.a.O., S. 10

und die außerdem für alle Personen gelten soll:

$$\forall (i) \{ Y_i = f(X_i) \}^1)$$

Nimmt man noch zusätzlich eine lineare Beziehung zwischen Lebensalter und IQ an, so haben die drei Arten von Hypothesen folgende Form:

$$(1) \quad \forall (i) \{ Y_i = a X_i + b \}$$

$$(2) \quad \forall (i) \{ Y_i = a_i X_i + b \} \quad \text{oder}$$

$$\forall (i) \{ Y_i = a_i X_i + b_i \}$$

$$(3) \quad \forall (i) \{ \bar{Y}_i = a X_i + b \}$$

Bei (1) und (2) werden die Individuen, bei (3) wird das Lebensalter durch den Alloperator gebunden. Nur bei (3) kann man unserer Meinung nach von einer Ausblendung des Individuums sprechen, "da der postulierte Zusammenhang zwischen beiden Variablen selbst dann für keine einzige Person zutreffen muß, wenn der Mittelwertstrend exakt linear ist"²⁾. Der globale Vorwurf, empirische Psychologie oder Pädagogik abstrahiere von der Individualität des Schülers, trifft deshalb nicht auf Hypothesen der Form (2) zu.

Für die empirische Forschung ergibt sich die forschungspraktische Konsequenz, daß im Extremfall eine Versuchsperson genügt, um die allgemeine Hypothese (1) zu falsifizieren. "Fragwürdig wird damit "nur eine bestimmte Variante psychologischer Forschung, die durch möglichst hohe Versuchspersonenzahlen, Verwendung statistischer Verfahren, relativ kurze Versuchsdauer und Desinteresse an der Individualität der einzelnen Versuchspersonen gekennzeichnet werden kann."³⁾

Fünf Bedingungen postulieren wir für den enger gefaßten Begriff 'Hypothese'⁴⁾:

1) Zur prädikatenlogischen Schreibweise:

KLAUS, G.: Moderne Logik. Abriß der formalen Logik, Berlin 1970²⁾, S. 236 ff. und SEGETH, W.: Elementare Logik, Berlin 1972⁷⁾, S. 121 ff.

2) BREDEKAMP, J., a.a.O., S. 14

3) vgl. dazu die WESTMEYER-HOLZKAMP-Kontroverse, in: WESTMEYER, H.: Kritik der psychologischen Unvernunft. Probleme der Psychologie als Wissenschaft, Stuttgart/Berlin/Köln/Mainz 1973, S. 129

4) WEINGARTNER, P.: Wissenschaftstheorie I. Einführung in die Hauptprobleme, Bad Cannstatt 1971, S. 57

1. Sie "bezieht sich (mindestens auch) auf bisher n o c h n i c h t u n t e r s u c h t e F a k t e n oder sogar auch auf unbeobachtbare Fakten".
2. Sie "ist auf Grund von neuem Wissen k o r r i g i e r b a r."
3. Sie hat eine "i m p l i k a t i v e F o r m".
4. Sie hat "eine gewisse U n i v e r s a l i t ä t".
5. Sie hat einen "G e h a l t".

Weniger plausibel als die Bedingungen 1. bis 4. erscheint die 5. Voraussetzung. Nach dieser Unterstellung sei der Begriff 'Hypothesengehalt' expliziert.

Hypothesen haben einen hohen Gehalt, wenn "eine Aussage uns relativ viel über die Realität sagt, d.h. daß sie vieles, was der Fall sein könnte, ausschließt, anders gesagt: sie ist mit vielem, was geschieht, nicht vereinbar."¹⁾

Veranschaulichen wir dies beispielhaft an einer Hypothese im Rahmen unserer Untersuchung. Dazu setzen wir fest:

P = def. Schüler x ist hochhängstlich

Q = def. Schüler x löst mit Erfolg hoch strukturierte Aufgaben.

Die Wahrheitstafel der Abbildung 5 zeigt die aus P und Q gebildeten komplexen Sätze mit unterschiedlichem Informationsgehalt. Die Konjunktion hat dabei den größten empirischen Gehalt, da sie die meisten Fälle ausschließt. "Der I n f o r m a t i o n s g e h a l t eines Satzes ist gleich der Anzahl der Bewertungen, bei denen der Satz f a l s c h ist."²⁾ Komplementär zum Gehalt

Abb. 5: Darstellung des Gehalts einer Hypothese in komplexen Sätzen

P	Q	$P \rightarrow Q$	$P \equiv Q$	$P \wedge Q$	$P \vee Q$
1		2	3	4	5
(1)W	W	W	W	W	W
(2)W	F	(F)	(F)	(F)	W
(3)F	W	W	(F)	(F)	W
(4)F	F	W	W	(F)	(F)

1) OPP, K.-D.: Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theoriebildung, Hamburg 1970, S. 57
 2) OPP, K.-D., a.a.O., S. 168

eines Satzes verhält sich dessen Spielraum: "Der Spielraum ist gleich der Anzahl der Bewertungen, bei denen ein Satz wahr ist."¹⁾ Die Begriffe Spielraum und Gehalt stellen somit auch Kriterien dar, die es erlauben, den Stellenwert von Hypothesen für die Theoriebildung anzugeben.

Weiter oben (S. 33) trennten wir in Anlehnung an METZGER nach den Funktionen, die ein Experiment haben kann, zwischen Erkundungs- und Entscheidungsversuch. Damit rechtfertigen wir die Unterscheidung zwischen 'offenen' und 'geschlossenen' Hypothesen, obwohl einige Bedenken vorgetragen werden.

In der Fachliteratur findet man "viele Experimente, die nicht von prognostischen Hypothesen ausgehen, bei denen das Ergebnis des Experiments also nicht vorhergesagt wird und somit auch nicht falsifizierbar ist. Die Hypothesen sind dann meist in Frageform formuliert, sie bezeichnen nur die zu untersuchenden Variablen, überlassen es aber dem Experiment, die Art des Kausalzusammenhanges zu spezifizieren. Wir wollen in diesem Falle von 'offenen' Hypothesen sprechen, gegenüber den 'geschlossenen' oder prognostischen Hypothesen bei einem Falsifizierungsversuch ... Der Unterschied der Hypothesenform beim Erkundungs- und Entscheidungsexperiment hat eine forschungstechnische Konsequenz. Denn der Offenheit und Breite der Untersuchungsaufgabe beim Erkundungsexperiment muß eine entsprechend breit angelegte Ergebnisermittlung folgen, da man ja zunächst noch nicht genau weiß, welche Wirkungen auftreten werden. Während beim Entscheidungsexperiment die Variablen, die eine Wirkung zeitigen sollen, durch die prognostische Hypothese spezifiziert sind, müssen beim Erkundungsexperiment in der Regel ein weites Variablenspektrum umfassende Beobachtungs- und Meßverfahren angesetzt werden, um eine große Zahl von potentiellen Wirkungen erfassen zu können."²⁾

1) OPP, K.-D., a.a.O., S. 168

2) SCHULZ, W.: Kausalität und Experiment in den Sozialwissenschaften. Methodologie und Forschungstechnik, München 1970, S. 117 - 118

Die Unterscheidung zwischen 'offenen' und 'geschlossenen' Hypothesen erscheint problematisch, wenn man OPPs Kritik der "Jagd nach abhängigen Variablen" akzeptiert. Diese Strategie kann so charakterisiert werden: "Man entdeckt ein Merkmal, das als Bestandteil der wenn-Komponente von Sätzen die Formulierung einer Vielzahl wahrer Gesetzesaussagen erlauben soll"¹⁾, die aber disjunkt sind:

- (1) Hochhängstliche Schüler (x) beurteilen im Anschluß an einen Englischtest (E) ihr Selbstbild negativ (S-).
- (2) Hochhängstliche Schüler haben nach einem Englischtest eine negative Einstellung zur englischen Sprache (EeS-).

Daraus bilden wir die logisch nicht gleichwertigen Hypothesen

$$H_1: \forall x \{ Ex \rightarrow S-x \wedge ES-x \}$$

$$H_2: \forall x \{ Ex \rightarrow S-x \vee ES-x \}$$

OPP gibt zwei Regeln an, bei deren Erfüllung die beschriebene Praktik gestattet ist: "E r s t e n s sollten bei der Jagd nach abhängigen Variablen Aussagen mit einer 'komplexen' dann-Komponente formuliert werden. Es sollten also nicht isolierte Aussagen ... gebildet, sondern die Merkmale der dann-Komponente sollten in einer einzigen dann-Komponente zusammengefaßt werden. Der Grund ist, wie wir sahen, daß andernfalls nicht klar ist, wie solche isolierten Hypothesen bei der Erklärung und Voraussage anzuwenden ist. Z w e i t e n s sollte die dann-Komponente möglichst eine Konjunktion sein."²⁾

4.2.2. Geschlossene Hypothesen

Entwicklung kann im Anschluß an WERNER³⁾ und LEWIN⁴⁾ als Differenzierungsvorgang aufgefaßt werden. "Dieses Gesetz der Differenzierung scheint in allen Bereichen der Entwicklung Geltung zu besitzen. So beobachten wir auf dem Sektor der Wahrnehmung eine zunehmende Differenzierung. Waren die Gegenstände ursprünglich in den Handlungsvollzug eingebettet, so erfolgt die spätere

1) OPP, K.-D., a.a.O., S. 189

2) OPP, K.-D., a.a.O., S. 192

3) WERNER, H.: Einführung in die Entwicklungspsychologie, München 1959/4

4) LEWIN, K.: Field Theory in Social Science, New York 1951

Dingerkenntnis demgegenüber unabhängig vom augenblicklichen Tätigsein. Weiterhin wird aus dem ungegliederten Wahrnehmungserlebnis ein strukturierter, gegliederter Sinneseindruck, aus der Verschmolzenheit der Wahrnehmung mit Gefühls- und Bedürfniszuständen eine vergleichsweise objektive Erfassung der Umwelt."¹⁾

In ihrem Literaturbericht zum Problem "Angst und Leistung" erklärt GARTNER-HARNACH die Untersuchungsergebnisse von MATSUDA & MATSUDA und Mc COY - anders als diese Autoren - mit der Differenzierungshypothese. "Alternativ ließe sich das genannte Ergebnis damit erklären, daß die SÄ (sehr ängstliche Schüler - K.H.) zu wenig die ihnen zur Verfügung gestellten Informationen verwerteten, so daß sie die beiden Situationen nicht unterscheiden konnten. Diese Interpretation leitet sich aus den theoretischen Formulierungen von EASTERBROOK (1959) zur mangelnden Signalnutzung Ängstlicher ab. Nichtausreichende Differenzierungsfähigkeit bildet möglicherweise auch die Ursache dafür, daß bestimmte Kinder sich ... Sorgen um die Schule machen, unabhängig davon, wie die Unterrichtsbedingungen gestaltet werden. DUNN (1968) argumentiert, ohne sich auf EASTERBROOK zu beziehen, ähnlich, nachdem er entgegen seiner Erwartung keine Interaktion zwischen Angstniveau und differentieller Instruktion fand. Er sieht die Möglichkeit, daß ein leistungsängstliches Kind seine Reaktionen nicht den situativen Gegebenheiten anpassen kann, sondern nur ein Alles - oder - Nicht - Verhaltensschema zur Verfügung hält."²⁾

-
- 1) OERTER, R.: Moderne Entwicklungspsychologie, Donauwörth 1973, S. 22
- 2) GARTNER-HARNACH, V.: Angst und Leistung, Weinheim/Basel 1972, S. 114. Sie bezieht sich auf die folgenden Arbeiten: MATSUDA, M./MATSUDA, F.: The effects of motivation levels and test anxiety levels on memory learning in children, in: Japanese J. Educ. Psychol., 1968; Mc COY, N.: Effects of test anxiety on children's performance as a function of instruction and type of task, in: J. pers. soc. Psychol., 1965; EASTERBROOK, J.A.: The effect of emotion on cue utilization and the organization of behavior, in: Psychol. Rev., 1959; DUNN, J.A.: Anxiety, stress and the performance of complex intellectual tasks: A new look at an old question, in: J. Consult. Clin. Psychol., 1968; DUNN, J.A.: The approach-avoidance paradigm as a model for the analysis of school anxiety, in: J. Educ. Psychol., 1968

Die generelle Hypothese, daß die zunehmende Differenzierungsfähigkeit eine Funktion des Alters ist, läßt vermuten, daß sich Schüler des 5. und 6. Schuljahres in den kognitiven und/oder emotionalen Leistungen unterscheiden:

- H (1): Schüler des 5. Schuljahres unterscheiden sich signifikant in der Beurteilung der Konzepte SELBSTBILD/WUNSCHBILD/SCHWACHER SCHÜLER/ENGLISCHLEHRER/KLASSENARBEIT/ZENSUR/KLASSENTEST von den Schülern des 6. Schuljahres.
- H (2): Schüler des 5. Schuljahres unterscheiden nicht zwischen den Konzepten KLASSENARBEIT und KLASSENTEST.
- H (3): Schüler des 6. Schuljahres unterscheiden zwischen den Konzepten KLASSENARBEIT und KLASSENTEST.

Geht man von der psychoanalytischen Annahme aus, daß ängstliche Kinder in der Prüfungssituation unbewußte Konflikte der Eltern-Kind-Beziehungen aufgrund der affinen Struktur von Lehrer- Eltern-Verhalten auf den Lehrer projizieren, kann man sich der Vermutung SARASONS anschließen: "1. Die Reaktion des prüfungsängstlichen Kindes auf Prüfungen und prüfungsähnliche Situationen in der Klasse spiegelt seine Erfahrungen in psychologisch oder interpersonell ähnlichen Situationen in seinem Zuhause wider; vor und nach dem Beginn der formellen Einschulung. 2. Die Prüfungssituation hat, zusätzlich zu ihrer bewußten Bedeutung, eine gleichzeitige unbewußte Bedeutung, die sich vornehmlich auf die Erfahrungen bezieht, die das Kind in der familiären Situation gemacht hat."¹⁾

Die Prüfinstrumente Klassenarbeit und Klassentest unterscheiden sich durch die Form der Aufgabendarbietung und durch die Art und Weise der Instruktion. Während die standardisierte Instruktion des Klassentests vom Tonband kommt, somit keine Rückfragen erlaubt, ist dies - wenn auch in geringem Maße - bei der Klassenarbeit möglich. Der Zeitdruck²⁾ ist bei der Klassenarbeit ebenfalls geringer, so daß die Prüfungssituation im ganzen entspannter ist.

1) SARASON, S.B. u.a.: Angst bei Schulkindern. Ein Forschungsbericht, Stuttgart 1971, S.

2) BARTMANN, Th.: Der Einfluß von Zeitdruck auf die Leistung und das Denkverhalten von Volksschülern, in: Psychol. Forschung, 1963
vgl. dazu die amerikanischen Literaturangaben bei GÄRTNER-HARNACH, V., a.a.O., S. 112

- H (4): Aufgrund der experimentellen Beeinflussung 'Klassentest' wird das Konzept UNSER ENGLISCHLEHRER von sehr ängstlichen Schülern negativ bewertet.
- H (5): Ein Prüfinstrument, das weniger 'autoritär' anmutet, wird von sehr ängstlichen Schülern positiv beurteilt.
- H (6): Die sehr ängstlichen Schüler des 6. Schuljahres werden das Konzept KLASSENTTEST negativer beurteilen als die Schüler des 5. Schuljahres.
- H (7): Die Hypothesen 4,5, und 6 gelten auch für das Konzept ZENSUR

Geht man von der weiter unten (S. 56) gemachten Annahme aus, daß die Beurteilungen der verschiedenen Konzepte die momentale Gefühlslage reflektieren, welche durch die unterschiedlichen Prüfungssituationen determiniert ist, dann erscheinen folgende Hypothesen zum Selbstbild sehr ängstlicher und wenig ängstlicher Schüler plausibel:

- H (8): Sehr ängstliche und wenig ängstliche Schüler unterscheiden sich in ihrem Selbstbild im Anschluß an verschiedene Prüfungssituationen.
- H (9): Wunschild und Selbstbild stimmen nach hoch strukturierten Prüfungssituationen mehr überein als nach wenig strukturierten.
- H (10): H (9) tritt schärfer bei Schülern des 6. Schuljahres hervor.

Die Messung der Einstellung der Schüler gegenüber England, den Engländern und der englischen Sprache mit der Einstellungsskala "Englisch" versucht, überdauernde Einstellungen zu erfassen. Sehr ängstliche und wenig ängstliche Schüler dürften sich deshalb, unabhängig von der momentanen Prüfungssituation, in ihrer Einstellung nicht signifikant unterscheiden. Ein experimenteller Einfluß dürfte nicht auftreten.

- H (11): Sehr ängstliche und wenig ängstliche Schüler unterscheiden sich nach den verschiedenen Prüfungsbedingungen Klassenarbeit und Klassentest nicht in ihrem Einstellungsscore.

Die Einstellungsskala "Englisch" übernimmt somit in diesem Experiment eine Kontrollfunktion.

4.2.2. Offene Hypothesen

Die Hypothesen H (1) - H (10) beziehen sich auf die Dimension Bewertung (B) des Semantischen Differentials. Über Beurteilungs-

tendenzen sehr ängstlicher und wenig ängstlicher Schüler, ihrem Beurteilungsverhalten gegenüber den Faktoren Aktivität (A) und Stärke (S), können wir keine aus dem wissenschaftlichen Kontext heraus begründeten Prognosen abgeben, so daß "das Ergebnis des Experiments also nicht vorhergesagt wird und somit auch nicht falsifizierbar ist. Die Hypothesen sind dann meist in Frageform formuliert, sie bezeichnen nur die zu untersuchenden Variablen, überlassen es aber dem Experiment, die Art des Kausalzusammenhangs zu spezifizieren."¹⁾

H (12): Welche emotionalen Reaktionen gegenüber den Konzepten SELBSTBILD/WUNSCHBILD/SCHWACHER SCHÜLER/GUTER SCHÜLER/ENGLISCHLEHRER/KLASSENARBEIT/ZENSUR/KLASSENTEST - gemessen mit den Skalen der Dimensionen Aktivität (A) und Stärke (S) des Semantischen Differentials - lösen die unterschiedlich konzipierten Prüfinstrumente normorientierter Leistungsmessung aus?

4.3. Der Versuchsplan

Die Bedenken DIETRICHs gegen eine experimentelle Organisation von Untersuchungseinheiten in Schulklassen seien vorangestellt: "Alle Klassen, insbesondere dann, wenn sie sich als pädagogische Gruppe organisiert haben, sind eben noch etwas anderes als Experimentalgruppen, nämlich in ihrer Leistungs- und Sozialstruktur äußerst differenzierte und störbare Gebilde. Werden sie experimentell manipuliert, besteht die Gefahr, daß man schiefe Ergebnisse erhält, freilich nun nicht infolge von Innen- und Außenfehlern, sondern infolge der Verletzung ihrer pädagogischen Eigenart ... Aber die Anwendbarkeit des exakten Experimentierens nimmt offensichtlich um so mehr ab, je komplexer und komplizierter die unabhängigen Variablen in ihrem inneren Aufbau, ihrem pädagogischen Anforderungsniveau und ihrer Handhabung sind und je weniger sie den Charakter einer bloßen 'Technik' besitzen."²⁾

Das spezifische Charakteristikum der traditionellen Prüfungssituation ist die restringierte Kommunikationsstruktur. Soziale

1) SCHULZ, W., a.a.O., S. 116

2) DIETRICH, G.: Bildungswirkungen des Gruppenunterrichts. Persönlichkeitsformende Bedeutung des gruppenunterrichtlichen Verfahrens, München 1969, S. 55

vgl. WINNEFELD, F.: Pädagogischer Kontakt und pädagogisches Feld, München/Basel 1963

Kontakte sind nicht gestattet, informelle soziale Konfigurationen werden außer kraft gesetzt, und die Integration der Leistungsmessung in den Lernprozeß fehlt in der Regel¹⁾. Wir haben es also mit einer artifiziellen Situation zu tun, für die ein experimenteller Versuchsplan maßgeschneidert werden kann.

Die Auswahl des Versuchsplans wird durch die Art des Problems gesteuert. Unsere Fragestellung, die Wechselwirkungen mindestens zweier Faktoren hypothetisch annimmt, kann deshalb methodisch nicht optimal mit einem einfachen Versuchsplan angegangen werden. "Die m o d e r n e V e r s u c h s p l a n u n g (experimental design) unterscheidet sich von dem klassischen oder traditionellen Vorgehen dadurch, daß stets mindestens 2 Faktoren zugleich untersucht werden ... Das Prinzip der modernen statistischen Versuchsplanung besteht darin, die Faktoren so einzusetzen, daß sich ihre E f f e k t e u n d W e c h s e l w i r k u n g e n sowie die Variabilität dieser Effekte m e s s e n, untereinander v e r g l e i c h e n und gegen die zufällige Variabilität a b g r e n z e n lassen."²⁾

Der Untersuchung liegt folgender unvollständige, zweifaktorielle Versuchsplan zugrunde:

Abb. 6: Zweifaktorieller Versuchsplan der Untersuchung

Prüfinstrumente				
	A ₁		A ₂	
Schüler- merkmal	B ₁	$\frac{N}{4}$ (a)	$\frac{N}{4}$ (b)	$\frac{N}{2}$
	B ₂	$\frac{N}{4}$ (c)	$\frac{N}{4}$ (d)	$\frac{N}{2}$
		$\frac{N}{2}$	$\frac{N}{2}$	N

Es bedeuten: A₁ = Informeller Test B₁ = Sehr ängstlicher Schüler
A₂ = Klassenarbeit B₂ = Wenig ängstlicher Sch.

1) METZGER, W.: Stimmung und Leistung. Die affektiven Grundlagen des Lernerfolgs, Münster 1967⁴, S. 2

2) SACHS, L.: Statistische Auswertungsmethoden, 1969², S. 539

Die unabhängige Variable 'Angst' - gemessen mit der TASC - hat zwei Stufen: Sehr ängstliche Schüler (SÄ) und wenig ängstliche Schüler (WÄ). Die Zuordnung der Schüler zu den Gruppen a,b,c oder d erfolgte durch Parallelisierung mit anschließender Randomisierung der Blöcke. Die Kriteriumsleistung X (die Kreuze im Fragebogen) wird als abhängig betrachtet von der Prüfungsmethode und dem Schülermerkmal 'Ängstlichkeit': $X = f(A,B)$.

"Ein solcher Plan ist immer dann angezeigt, wenn folgende Hypothesen überprüft werden sollen:

- daß die Methode die Kriteriumsleistung beeinflusst,
- daß das Schülermerkmal die Kriteriumsleistung beeinflusst,
- daß der Methodeneffekt beeinflusst wird von der Ausprägung des Schülermerkmals."¹⁾

Außerdem ist zu überprüfen, ob die Grundprinzipien der Versuchsplanung²⁾ eingehalten worden sind: 1. W i e d e r h o l u n g, 2. R a n d o m i s i e r u n g und 3. B l o c k b i l d u n g.

LINDER formuliert diese Prinzipien so:

"Grundsatz I. Um den Versuchsfehler zuverlässig bestimmen zu können, müssen die Verfahren über mehrere Versuchseinheiten wiederholt werden."

"Grundsatz II. Damit jede Einseitigkeit in der Zuteilung der Versuchseinheiten auf die verschiedenen Verfahren nach Möglichkeit vermieden wird, muß diese Zuteilung zufällig vor sich gehen."

"Grundsatz III. Um die Empfindlichkeit des Versuchs zu erhöhen, und um gleichzeitig eine möglichst breite induktive Basis zu gewährleisten, werden Blöcke mit unter sich möglichst gleichartigen Versuchseinheiten gebildet."³⁾

Die randomisierte Zuteilung der Versuchseinheiten auf die einzelnen Blöcke ist "empfindlicher als der vollständig randomisierte Plan"⁴⁾.

Ein Versuchsplan ist daraufhin zu analysieren, wieweit er die Kriterien der inneren und äußeren Gültigkeit berücksichtigt.

1) KLAUER, K.J.: Das Experiment in der pädagogischen Forschung. Eine Einführung, Düsseldorf 1973, S. 120

2) SACHS, L., a.a.O., S. 537

3) LINDER, A.: Planen und Auswerten von Versuchen. Eine Einführung für Naturwissenschaftler, Mediziner und Ingenieure, Basel / Stuttgart 1959², S. 10, 12 und 14

4) SACHS, L., a.a.O., S. 541

"I n n e r e G ü l t i g k e i t ist die minimale Voraussetzung, ohne die kein Experiment interpretierbar ist: Verursachen die experimentellen Einwirkungen (Behandlungen) wirklich einen Unterschied in dieser besonderen experimentellen Situation? Ä u ß e r e G ü l t i g k e i t wirft die Frage der G e n e r a l i s i e r b a r k e i t auf: Für welche Populationen, Gegebenheiten, Behandlungs- und Meßvariablen kann dieser Effekt generalisiert werden?"¹⁾

Es sind acht Störvariablen, deren Varianzen in die Effekte der unabhängigen experimentellen Variablen eingehen können: Zwischenzeitliches Geschehen, Reifung, Testen, Hilfsmittel, statistische Regression, Verzerrungen (biases), experimentelle Einbußen, Wechselwirkung zwischen Auswahl und Reifung. Die Kontrolle dieser acht Faktoren durch das Design erscheint plausibel und evident. "Doch diese Aussage kann nur mit Vorbehalt akzeptiert werden. Die von CAMPBELL und STANLEY behandelten Faktoren - im Grunde gleichzeitig Validitätskriterien - können nicht als Einzelfaktor isoliert und kontrolliert werden, denn sie sind sehr komplexe Faktorengruppen, innerhalb derer und zwischen denen Interdependenzen bestehen dürften."²⁾

Grundsätzlich problematisch erscheint die Aufspaltung des Variablenatzes in 'unabhängige' und 'abhängige' Variablen. ROTH argumentiert: "Die scheinbar 'unabhängige Variable', als deren Funktion die 'abhängige Variable' durch das Experiment dargestellt werden soll, erweist sich somit als Punkt in einem eng verflochtenen System wechselseitiger Abhängigkeiten, dessen angeblich isolierte Variation im klassischen Experiment die tatsächliche Funktion im Bedingungssatz menschlichen Verhaltens nicht adäquat trifft."³⁾

-
- 1) SCHWARZ, E.: Experimentelle und quasiexperimentelle Anordnungen in der Unterrichtsforschung, in: Handbuch der Unterrichtsforschung, Teil I, Weinheim 1970, S. 459
 - 2) ROTH, L.: Effektivität von Unterrichtsmethoden. Empirische Untersuchungen zu Wirkungen der Organisationsformen von Lernbedingungen, Hannover/Berlin/Darmstadt/Dortmund 1971, S. 35
 - 3) ROTH, E.: Persönlichkeitspsychologie. Eine Einführung, Stuttgart/Berlin/Köln/Mainz 1969, S. 24

Die vorgelegte Untersuchung ist eine Pilotstudie, deren Ergebnisse nicht generalisierbar sind. Die Kontrolle der Variablen, welche die externe Validität ausmachen, erfordert erheblichen finanziellen und personellen Aufwand, der für eine Dissertation in der Regel nicht zur Verfügung steht. Die Auswertung der Daten mit Hilfe statistischer Signifikanztests erscheint gerechtfertigt, da "eine Möglichkeit zur legitimen Verwendung statistischer Verfahren darin gesehen wird, diese als approximative randomization tests interpretieren zu können, welche die Annahme der Zufallsstichprobenentnahme nicht nötig machen" und "weil statistisch nicht auf Personen geschlossen werden solle, die am Experiment nicht teilgenommen haben; der statistische Schluß eines randomization tests beziehe sich auf die experimentellen Effekte"¹⁾. Die Varianz der abhängigen Variablen wird also zu Lasten der experimentellen Bedingungen interpretiert. Die Funktion unseres Experiments besteht darin, Vorstufe für weitere Forschung auf diesem Gebiet zu sein.

4.4. Die experimentellen Gruppen

Für das Experiment standen vier Klassen des 5. und vier Klassen des 6. Schuljahres der Städtischen Hauptschule Rütten zur Verfügung. Es handelt sich um eine Vollerhebung zweier Jahrgänge einer Schule. Da nur der Nachweis des Effekts bestimmter experimenteller Bedingungen beabsichtigt wurde, nicht jedoch repräsentative Aussagen über bestimmte Merkmalszusammenhänge in einer definierten Gesamtheit gewonnen werden sollten, entfiel das Problem, faktorengleiche Stichproben zu konstruieren. Das Problem reduziert sich damit auf die Zusammenstellung faktorengleicher Parallelgruppen für bestimmte experimentelle Bedingungen. "Repräsentativität wird insbesondere gefordert, wenn man Aussagen über bestimmte Merkmalsverteilungen in der Gesamtheit machen will, jedoch nicht unbedingt, wenn man Aussagen über vermutliche Kausalzusammenhänge

1) BREDEKAMP, J.: Der Signifikanztest in der psychologischen Forschung, Frankfurt am Main 1972, S. 30

unter konstanten Bedingungen (d.h. ceteris paribus) machen will."¹⁾

Der Entwurf experimenteller Bedingungen antizipiert hypothetische Zusammenhänge aufgrund theoretischer Überlegungen. "Solche Aussagen lassen sich oft schon durch sehr kleine Stichproben zumindest in ihrer Allgemeingültigkeit als unhaltbar widerlegen. Die Stichproben brauchen dabei natürlich auch nicht repräsentativ für eine Grundgesamtheit zu sein."²⁾

Um die Ausgangsbedingungen zu egalisieren, mußten faktorengleiche Parallelgruppen gebildet werden. Als relevanteste unabhängige Variable, die kontrolliert werden mußte, hatte sich nach der Literaturdurchsicht die Variable 'Angst' herauskristallisiert, speziell die Variable 'Prüfungsangst'.

Der Ausprägungsgrad des Merkmals 'Angst' eines jeden Schülers wurde deshalb mit den von NICKEL und SCHLÜTER³⁾ auf deutsche verhältnisse übertragenen und teilweise modifizierten Angstskalen von CASTANEDA u.a. (CMAS) und SARASON u.a. (TASC) gemessen. Dann wurde von den Schülern des 5. Schuljahres eine Rangreihe nach dem Kriterium 'Prüfungsangst' gebildet. Nach dem Zufall wurde entschieden, welcher von zwei Schülern mit aufeinanderfolgendem Rangplatz der experimentellen Bedingung 'Klassenarbeit' oder 'Klassentest' angehören sollte. Somit haben wir es mit einer unvollständigen Randomisierung zu tun. "Unter (vollständiger - K.H.) Randomisierung versteht man eine rein zufällige Gruppenbildung."⁴⁾

-
- 1) und 2) MAYNTZ, R./HOLM, K./HUBNER, P.: Einführung in die Methoden der empirischen Soziologie, Köln/Opladen 1969, S. 69
3) NICKEL, H./SCHLÜTER, P.: Angstwerte bei Hauptschülern und ihr Zusammenhang mit Leistungs- sowie Verhaltensmerkmalen, Lehrerurteil und unterrichtsstil, in: Zeitschr. f. Entw.-ps. und Pädag. Ps., 1970

NICKEL, H./SCHLÜTER, P./FENNER, H.-J.: Angstwerte, Intelligenztest- und Schulleistungen sowie der Einfluß der Lehrerpersönlichkeit bei Schülern verschiedener Schularten, in: Psychologie in Erziehung und Unterricht, 1973, S. 10 - 12

- 4) KLAUER, K.J., a.a.O., S. 115

Tabelle 1: Aufteilung der Versuchsgruppen

	5. Schuljahr	6. Schuljahr
Klassenarbeit	$\frac{S\bar{A}}{W\bar{A}}$ 25	$\frac{S\bar{A}}{W\bar{A}}$ 25
	25	25
Klassentest	$\frac{S\bar{A}}{W\bar{A}}$ 25	$\frac{S\bar{A}}{W\bar{A}}$ 25
	25	25
	$N_1=100$	$N_2=100$

Außer der Variablen 'Angst' wurden noch das Alter, die Fremdspracheneignung und der Intelligenzquotient eines jeden Schülers erfaßt.

Mit zwei Prüfverfahren aus der inferenten Statistik, dem t-Test und dem χ^2 -Test, wurde kontrolliert, ob sich die Stichproben nur innerhalb eines vorher festgelegten Zufallsbereichs unterschieden.

Das Skalenniveau der zusammengefaßten TASC/MASC-Werte entspricht einer Nominalskala, für welche die Postulate der Identität¹⁾ gelten: 1. $a = b$ oder $b = a$, 2. wenn $a = b$, dann $b = a$ und 3. wenn $a = b$, $b = c$, dann $a = c$.

Wir rechneten nach der Formel:

$$\chi^2 = \frac{(ad - bc)^2 \cdot n}{(a+b)(c+d)(a+c)(b+d)} \quad 2)$$

Schon die grafische Darstellung wies auf einen nicht signifikanten Unterschied hin.

Auch für die Versuchsgruppen des 6. Schuljahres konnte keine Signifikanz festgestellt werden, wie das aus der Tabelle 2 errechnete χ^2 zeigt.

1) SIXTL, F.: Meßmethoden der Psychologie, Weinheim 1967, S. 4

2) CLAUSS, G./EBNER, H.: Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen, Berlin 1968, S. 238

Grafik 1: TASC/MASC

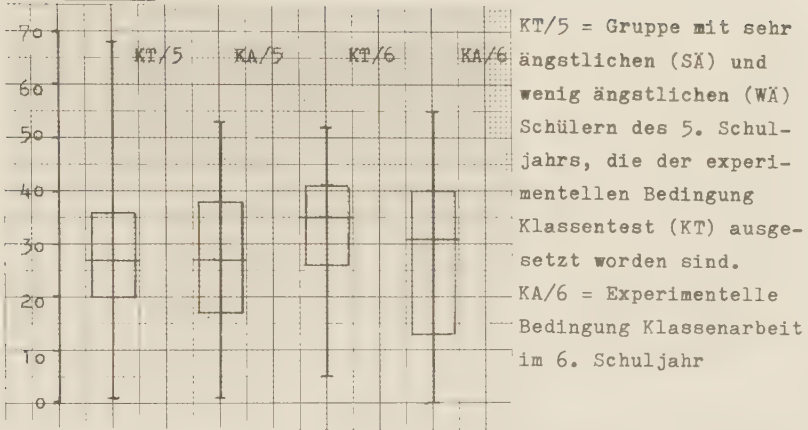


Tabelle 2: Ängstlichkeitsgrad und experimentelle Bedingung Klassentest vs. Klassenarbeit im 5. Schuljahr

	KT/5	KA/5	
WÄ: $X = Z_{50}$	26	26	52
SÄ: $X = Z_{50}$	24	24	48
	50	50	100 $\chi^2 = 0.04$ ns.

Tabelle 3: Ängstlichkeitsgrad und experimentelle Bedingung Klassentest vs. Klassenarbeit im 6. Schuljahr

	KT/6	KA/6	
WÄ: $= Z_{50}$	23	27	50
SÄ: $= Z_{50}$	27	23	50
	50	50	100 $\chi^2 = 1.00$ ns.

Die Variablen Intelligenz, Fremdspracheneignung und Alter haben Intervallskalenniveau, so daß die Nullhypothese ($H_0: M_1 = M_2$), es bestehe kein signifikanter Unterschied zwischen den Mittelwerten der Gruppen KT/5 und KA/5 oder KT/6 und KA/6, mit einem

anspruchsvollen parametrischen Prüfverfahren getestet werden kann. In unserem Fall ist das der t-Test¹⁾, da seine Modellprämissen erfüllt werden können: Intervallskala, Normalverteilung der Variablen und gleiche Varianz der Stichproben.

Vor der Anwendung des t-Tests wurde die Homogenität der Varianzen überprüft. Die Prüfgröße ist $F = s_1^2/s_2^2$. Ist F kleiner als der Tabellenwert, wird H_0 angenommen; H_0 wird jedoch zurückgewiesen, wenn F gleich oder größer als der Tabellenwert ist:

Wenn $F < F_{0.05; n_1-1/n_2-1}$, dann wird H_0 angenommen,

wenn $F \geq F_{0.05; n_1-1/n_2-1}$, dann wird H_0 zurückgewiesen²⁾

"Diese F-Verteilung hat zwei Parameter, die Anzahl der Freiheitsgrade f_1 und f_2 . Für den Vergleich der Varianzen zweier unabhängiger Stichproben ist $f_1=n_1-1$ und $f_2=n_2-1$."³⁾

Tabelle 4: Homogenitätsprüfung der Stichprobenvarianzen bei den Variablen Intelligenz (LPS), Fremdspracheneignung (FTU) und Alter der Versuchsgruppen der Klassen 5 und 6

	s_1	s_2	s_1^2/s_2^2	F	$F_{0.05; n_1-1/n_2-1}$
LPS: 5. Sch.	9,421	9,308	1,02	1,02	1,97; 48/49 ns.
6. Sch.	10,683	8,794	1,48	1,48	1,98; 47/48 ns.
FTU 5. Sch.	7,566	6,752	1,26	1,120	1,96; 49/49 ns.
6. Sch.	8,778	6,652	1,74	1,319	1,97; 48/49 ns.
Al- 5. Sch.	7,066	7,064	1,00	1,000	1,91; 49/49 ns.
ter 6. Sch.	6,828	5,059	1,82	1,349	1,96; 49/49 ns.

Die Mittelwerte wurden nach der Formel

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \quad \text{auf Signifikanz untersucht.}$$

$$df = n_1 + n_2 - 2$$

Aus den Stichprobenstreuungen s_1 und s_2 wurde die Streuung s der Grundgesamtheit geschätzt:

1) CLAUSS, G./EBNER, H., a.a.O., S. 184 ff.

2) CLAUSS, G./EBNER, H., a.a.O., S.191

3) CLAUSS, G./EBNER, H., a.a.O., S. 191

$$s = \sqrt{\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}}$$

Als Kriterium für die Annahme der Nullhypothese gilt:

$$|t| < t_{0.05;df}, \text{ wobei } df = n_1 + n_2 - 2.$$

Als Kriterium für die Zurückweisung der Nullhypothese gilt:

$$|t| \geq t_{0.05;df}, \text{ wobei } df = n_1 + n_2 - 2.$$

Alle berechneten t-Werte der Tabelle 4 sind kleiner als die kritischen Werte und fallen daher in den zufälligen Bereich, so daß die Unterschiede zwischen den Stichprobenmittelwerten nicht signifikant sind.

Tabelle 5: Überprüfung der Nullhypothese der Stichprobenmittelwerte der je zwei Versuchsgruppen aus dem 5. und 6. Schuljahr bei den Variablen Intelligenz (LPS), Fremdspracheneignung (FTU) und Alter

		5. Schuljahr		6. Schuljahr	
		KT/5	KA/5	KT/6	KA/6
LPS:	$\bar{X}$	103,89	105,28	104,93	104,48
	s_1^2	88,760		114,145	
	s_2^2		86,654		77,338
	s	9,364		9,774	
	df	97		95	
	t	0,736	ns.	0,226	ns.
FTU:	$\bar{X}$	43,74	42,56	38,36	36,54
	s_1^2	57,257		77,070	
	s_2^2		45,598		44,253
	s	7,171		7,777	
	df	98		97	
	t	0,822	ns.	1,164	ns.
Alter:	$\bar{X}$	129,62	128,10	128,84	128,10
	s_1^2	49,913		46,626	
	s_2^2		49,928		25,602
	s	7,065		6,009	
	df	98		98	
	t	1,075	ns.	0,615	ns.

Wir konnten mit Methoden der schließenden Statistik nachweisen, daß sich die Versuchsgruppen des 5. Schuljahres und die des 6.

Schuljahres in den für relevant gehaltenen Variablen Angst, Alter, Intelligenz und Spracheignung in den Ausgangsbedingungen nur im Bereich zufälliger Abweichungen unterscheiden.

4.5. Das Untersuchungsinstrumentarium

Auffällig an der empirischen Unterrichtsforschung ist ihre Vorliebe für Effektivitätsuntersuchungen, die sich auf das Lernprodukt kognitiver Prozesse richtet, also "Produktivität, Leistung im engeren Sinn (Umfang des bewältigten Stoffes, Problemtiefe, Geschwindigkeit des Problemlösens, Leistungsdurchschnitt in Noten ausgedrückt)"¹⁾, erfassen. Abstrahiert verhält sich die Forschung gegenüber Effektivitätskriterien wie "Zufriedenheit der Mitglieder (Enge des sozialen Kontaktes, Klima, Zusammenhalt)"²⁾ oder " 'Emanzipation' der Gruppenmitglieder (Einstellungsänderungen, Zuwachs an Unabhängigkeit, Selbstreflexion und Selbstbestimmung, Persönlichkeitsentwicklung zum kritischen Denken hin)." ³⁾ Werden soziale Konfigurationen in der Schulklasse untersucht, so interessiert das Wissen um das soziale Geflecht lediglich um der Effektivitätssteigerung willen. Die Gruppe wird zum Mittel, der Sinn der informellen Gruppenstruktur pervertiert. "Die Frage nach der Leistung des informellen Systems zur Erfüllung offizieller Aufgaben ist fast ausschließlich vom Effektivitätsinteresse geleitet."⁴⁾

Erklärung für diese Interessenverlagerung der empirischen Unterrichtsforschung scheinen die gesellschaftlichen Rahmenbedingungen zu bieten, die sich im bürokratischen Verwaltungscharakter der Schule manifestieren, der "die Rolle einer bürokratischen Zuteilungsapparatur von Lebens-Chancen"⁵⁾ hat. Die formelle bürokratische Struktur des Schulsystems affiziert die informellen affektiven Gruppenbeziehungen, "deren Funktion sich nicht nach den Bedürfnissen der Gruppenmitglieder, sondern nach denen der

1) ULICH, D.: **Gruppendynamik** in der Schulklasse. Möglichkeiten und Grenzen sozialwissenschaftlicher Analysen, München 1971, S. 69

2) und 3) ULICH, D., a.a.O., S. 69

4) ULICH, D., a.a.O., S. 43

5) SCHELSKY, H.: **Soziologische Bemerkungen zur Rolle der Schule in unserer Gesellschaftsverfassung**. Eine Denkschrift (1956), in: Auf der Suche nach der Wirklichkeit, Düsseldorf/Köln 1965, S. 137

übergeordneten Institution richtet. Die institutionellen Grundlagen hierfür sind die Lehrpläne, Verordnungen, Richtlinien über Sitzordnungen, Didaktik, Prüfungs- und Zensurwesen, Unterrichts- und Erziehungsziele usw."¹⁾ Leistungszuwachsuntersuchungen scheinen eher im Interesse der 'übergeordneten Institution' zu sein als Untersuchungen, welche z.B. die Nebenwirkungen bestimmter organisatorischer Strukturen auf die affektiven Komponenten der Schülerpsyche aufdecken wollen. Es sollte nicht vergessen werden, "daß nämlich Erziehung insgesamt mindestens ebensosehr Bildung von Einstellungen wie Vermittlung von Wissen darstellt."²⁾

Unsere Vermutungen werden gestützt durch den Mangel an Meßinstrumenten, die pädagogische Nebenwirkungen zu erfassen erlauben. "Wo sich der Forscher jedoch für andere Erziehungsergebnisse als den Zuwachs an faktischem Wissen interessiert, sieht er sich häufig genötigt, sein eigenes Meßinstrument zu entwickeln."³⁾

Da unser Forschungsinteresse auf das durch normorientierte Prüfungsbedingungen ausgelöste emotionale Erleben des Schülers zielte, mußte ein Meßinstrument konstruiert und evaluiert werden, das die Einstellung des Schülers zu England, den Engländern und der englischen Sprache zu erfassen ermöglichte. Die Einstellung gegenüber dem englischen Sprachraum wollten wir deshalb erfassen, weil "die affektive oder Gefühlskomponente ... bei der Einstellung eine zentrale Bedeutung"⁴⁾ hat.

Das Ergebnis unserer Bemühungen ist die Einstellungsskala 'Englisch! Sie wurde nach einem psychometrischen Ansatz⁵⁾ konstruiert.

Ebenso wie sich der übergeordnete institutionelle Rahmen bei der Einstellungsbildung langfristig auswirkt, so werden auch bestimmte Formen der Leistungsüberprüfung⁶⁾ bestimmte Einstellungen der

1) ULICH, D., a.a.O., S 41

2) ROTH, E.: Einstellung als Determination individuellen Verhaltens, Göttingen 1967, S. 23

3) BRIM, G.O.: Soziologie des Erziehungswesens, Heidelberg 1963, S. 62

4) SULLWOLD, F.: Theorie und Methodik der Einstellungsmessung, in: Graumann, C.F. (Hrsg.): Hdb. d. Ps., Band 7/1, Halbband, Sozialpsychologie, Göttingen 1969, S. 476

5) Zur Kritik s. ROTH, E., a.a.O., S. 33

6) ROEDER, P.M.: Die pädagogische Bedeutung von Erfolgskontrolle und Leistungsmessung, in: Speck, J. (Hrsg.): Leistung, Erfolg und Erfolgskontrolle in der Pädagogik und ihren Nachbarwissenschaften, Münster 1968, S. 106 ff.

Schüler ergeben. Einstellungen sind "dem Verhalten zugrunde liegende Organisationsformen, die sich aus einem gegebenen Organismus in einer individuellen Lerngeschichte entwickeln und durch Gegenstandsbezug und Systemcharakter gekennzeichnet sind, vermögen alle diejenigen Gegebenheiten einheitlich im Bedingungssatz menschlichen Verhaltens in sich zu schließen, zu dessen Erklärung sie notwendigerweise vorausgesetzt werden müssen"¹⁾).

Einstellungen werden gelernt. Ihre Perseveration könnte als Funktion der Häufigkeit der Wiederholung bestimmter Erfahrungen bezeichnet werden. "Die Dauer der Verfügbarkeit oder der Wirksamkeit von etwas Erlerntem hängt also unter somit vergleichbaren Bedingungen von der Häufigkeit der Situation ab, in der etwas Spezifisches gelernt wurde ... Eine einmalige Erfahrung wird ebenso eine vorübergehende oder momentane Einstellung nach sich ziehen, wie etwa die Versuchsanweisung in einem der früheren Einstellungsexperimente ... Verhaltensbeeinflussend wirken nicht nur habituelle, sondern auch momentane Einstellungen, wie jeder an sich selbst beobachten kann und wie es auch experimentell erwiesen ist. Im Schema sieht diese Differenzierung wie folgt aus:

H ä u f i g k e i t d e r W i e d e r h o l u n g e i n s t e l -
l u n g s b i l d e n d e r E r f a h r u n g e n

gering

hoch

momentan

E i n s t e l l u n g

habituell"²⁾

Die geringe Testerfahrung der Schüler könnte eine momentane Einstellungsänderung zur Folge haben. Diese Momentaufnahme affektiver Prozesse ermöglicht das Semantische Differential, das die konotative Komponente von Begriffen erfaßt. Die nur kurz andauernde Prüfungssituation dürfte sich nicht auf habitualisierte Einstellungskomplexe wie z.B. das Stereotyp 'England' auswirken. Das Semantische Differential ist als methodisches Instrument im Gegensatz zur Einstellungsskala 'Englisch' sensibler gegenüber

1) ROTH, E., a.a.O., S. 145

2) ROTH, E., a.a.O., S. 76

nur kurz andauernden einstellungsbildenden Erfahrungen. Die 'Trägheit' der Einstellungsskala wirkt jedoch - methodisch gesehen - als Korrekturfaktor, da sie aufgrund der item-Auswahl resistent gegenüber momentanen Einflüssen ist. Signifikante Unterschiede zwischen den experimentellen Gruppen - diese Hypothese sei vorweggenommen - dürften sich in der Einstellung gegenüber England nicht ergeben.

4.5.1. Einstellungsskala 'Englisch'

Im deutschsprachigen Raum hat unseres Wissens nach nur HEUER versucht, "Vorurteile und Stereotypen im Englandbild der Schüler"¹⁾ mit Hilfe eines Fragebogens zu erfassen. Das vorgelegte Meßinstrument und die Anlage der einzelnen Untersuchungen entsprechen aber nicht den technischen Standards empirischer Forschungsmethoden. Aussagen wie z.B. "die direkte Proportionalität von Motivation und Lernerfolg hat sich deutlich abgezeichnet"²⁾ haben abgesehen von ihrer wissenschaftstheoretischen Problematik - wenig Wert, da das Meßinstrument Mängel im Konstruktions- und Evaluationsbereich aufweist. Methodologische Bewußtheit in der Konstruktionsphase eines Fragebogens oder einer Einstellungsskala zwingt zur Berücksichtigung des Bedeutungsaspekts. Ein Fragebogenitem besitzt für verschiedene Bezugsgruppen unterschiedliche Bedeutung, und "die Bedeutung der Reihe von Aussagen, die einen Fragebogen konstituiert, nimmt für jedes Sample von Versuchspersonen einen variablen Status an, wenn der Forscher sich nicht eine Theorie von Bedeutung und Rollenübernahme aneignet, die dem mechanischen oder deterministischen Gebrauch von Fragebogen mit vorfixierten Auswahlfragen entspricht."³⁾

BERGER ist in seiner Kritik noch schärfer als CICOUREL: "Die Ermittlung des praktischen Bedeutungskreises sprachlicher Ausdrücke führt also zu grundlegend anderen Ergebnissen als eine inhaltliche

1) HEUER, H.: England in Vorstellung und Vor-Wissen unserer Schüler. Die Ergebnisse einer empirischen Untersuchung, in: Englisch, 1966, S. 71 und HEUER, H.: Zur Motivation im Englischunterricht. Die Ergebnisse einer empirischen Untersuchung, in: Englisch 1968, S. 33 ff.

2) HEUER, H. (1968), a.a.O., S. 36

3) CICOUREL, A. von: Methode und Messung in der Soziologie, Frankfurt am Main 1970, S. 199
vgl. HABERMAS, J.: Zur Logik der Sozialwissenschaften, Frankfurt am Main 1970, S. 144

Bedeutungsanalyse." ¹⁾ Die einseitige Verlagerung des Forschungsinteresses - unter Absehung von speziellen Situationsumständen - auf die inhaltliche Bedeutungsanalyse habe letztlich ihren Grund im Rationale empirischer Methoden und in der Zielsetzung sozialwissenschaftlicher Methodologie, "Beobachtungsverfahren mit möglichst eindeutigen, präzisen und situationsunabhängigen Sprachregeln zu bevorzugen. Nur so können die Ansprüche möglichst strenger intersubjektivität und zuverlässiger Erkenntnis erfüllt werden. Seinen methodischen Niederschlag findet die Forderung in der Aufstellung allgemeiner Beobachtungsschemata, der Anwendung standardisierter Fragebögen mit festen Frage- Antwort-Kategorien, der Bildung von Meßsystemen und Skalen, die nicht nach Situationsumständen und gesellschaftlicher Lebensform (z.B. Klassenlage) differenzieren. Der Präzisionsanspruch hat Konsequenzen in der Einschränkung auf solche Beobachtungsergebnisse, die durch möglichst festliegende und situationsunabhängige Merkmale ausgezeichnet sind. Am deutlichsten wird das in der Einstellungs- und Meinungsforschung, die trotz gegenteiliger Untersuchungsergebnisse am Ideal eindeutiger und genereller, nicht nach Situationsumständen variierender Einstellungen festhält. Diese Fixierung kann methodisch aus den gewählten Beobachtungsverfahren erklärt werden, die wegen ihrer Standardisierung nur die Ermittlung unflexibler Verhaltensmuster zulassen." ²⁾

Aber gerade diese verfestigten oder stereotypen Einstellungen wollen wir mit unserer Einstellungsskala erfassen, so daß die von BERGER vorgetragene Kritik für unseren Untersuchungszweck geradezu ins Positive umschlägt: Die Konstruktionsbedingungen der Einstellungsskala 'Englisch' sind der Zielsetzung unserer Untersuchung affin.

4.5.1.1. Allgemeine methodologische Bemerkungen zur Einstellungsuntersuchung

Im Gegensatz zur Psychophysik fehlt bei der Einstellungsuntersuchung das physische Korrelat. Da jedoch das manifeste Verhal-

1) BERGER, H.: Erfahrung und Gesellschaftsform. Methodologische Probleme wissenschaftlicher Beobachtung, Stuttgart/Berlin/Köln/Mainz 1972, S. 144

2) BERGER, H., a.a.O., S. 143

ten eines Individuums in bezug auf ein bestimmtes 'psychologisches Objekt' eine gewisse Konsistenz zeigt, können wir hypothetisch auf sogenannte 'latente' Variablen schließen, die nun an die Stelle des physischen Korrelats treten. Dabei wird vorausgesetzt, daß die Stichprobe von 'statements' (Aussagen) für die zu untersuchende Dimension repräsentativ ist. Hier treten hermeneutische Probleme auf, da exakte Kriterien, Zuordnungsregeln, fehlen, die die Validität der Indikatoren als gelungene Konstruktoperationalisierungen garantieren. "So ist es auch unmöglich, theoretische Aussagen endgültig zu falsifizieren, weil eine solche Falsifikation die valide Messung der in den Hypothesen verwendeten Konstrukte voraussetzen würde." ¹⁾

Ziel der Einstellungsuntersuchung ist es, den Ort eines oder mehrerer Individuen auf dem jeweiligen psychologischen Kontinuum anzugeben. Dazu müssen die Behauptungen des Einstellungsfragebogens skaliert werden. "Skalierung der Aussagen bedeutet, daß der Ort jeder für die Einstellungsuntersuchung in Betracht gezogenen Aussage auf dem psychologischen Intensitätskontinuum festgestellt wird. Der Ort wird durch den sogenannten 'Skalenwert' der Aussage angezeigt." ²⁾

Bei der weiter unten beschriebenen Skalierungsmethode des Paarvergleichs wird vorausgesetzt, "daß die jeweilige Zustimmung oder Ablehnung eine Funktion der Intensität des positiven oder negativen Gefühls ist, das der Proband gegenüber dem gesamten 'psychologischen Objekt' besitzt." ³⁾ Damit stellt sich das Problem, wie weit die Skalierung der Aussagen von der Einstellung der jeweiligen Expertengruppe abhängt.

4.5.1.2. Skalierungsmethode 'Paarvergleich'

"THURSTONE .. hat gezeigt, daß es möglich ist, aus einer Verteilung von Alternativurteilen auf den subjektiven Größenunterschied zu schließen, den Paare von Objekten bezüglich eines gemeinsamen

1) STOSBERG, M.: Analyse der Massenkommunikation. Einstellungs-
messung, Düsseldorf 1972, S. 19

2) SÜLLWOLD, F., a.a.O., S. 480

3) SÜLLWOLD, F., a.a.O., S. 484

Merkmals für eine Vp haben. Die Bedingungen, unter denen dies möglich ist, und die Überlegungen, die diesen Schluß ermöglichen, bilden zusammen das Law of Comparative Judgement¹⁾). Die Bedeutung des L. of C.J. besteht darin, daß es ein Rationale für die Anordnung von Stimuli auf einem psychologischen Kontinuum gibt.²⁾

Die Versuchsform der Einstellungsskala 'Englisch' bestand aus vierzehn Behauptungen:

1. Ich habe nichts gegen Engländer, doch glaube ich, daß sie ein wenig komisch sind.
2. Meine Ferien würde ich schon in England verbringen, doch möchte ich nicht für immer dort wohnen.
3. Mir genügte es, einen englischen Brieffreund (eine englische Brieffreundin) zu haben.
4. Ich kann mir gut vorstellen, mit einem Engländer (einer Engländerin) verheiratet zu sein.
5. Anstelle von Englisch sollte man heutzutage mehr Mathematik unterrichten.
6. So sehr interessieren mich die Engländer, England und die englische Sprache auch wieder nicht, daß ich mich mit ihnen mehrere Jahre in der Schule beschäftigen möchte.
7. Englisch ist eine Weltsprache, so daß es sich lohnt, sie zu erlernen.
8. Ich möchte die englische Sprache am liebsten in England lernen.
9. Ich weiß zu wenig über England und die Engländer, um eine eigene Meinung zu haben.
10. Ich kann Engländer nicht leiden.
11. Ich kümmere mich überhaupt nicht um das, was mit England usw. zu tun hat.
12. Meine Einstellung gegenüber England, den Engländern und der englischen Sprache ist eher gleichgültig.
13. Das Erlernen der englischen Sprache weitet den Blick des Menschen.
14. Engländer sind fair.

Unter der Annahme, daß jede Behauptung (jeder Stimulus) auf der Merkmalsdimension 'anglophil' liegt, können wir postulieren, daß jede Behauptung eine bestimmte Merkmalsausprägung besitzt.³⁾

Der Grad der Ausprägung eines jeden Stimulus ist unbekannt und variiert bei jeder Behauptung. Ziel des Paarvergleichs ist es, den $n = 14$ dichotomen non-monotonen items⁴⁾ einen Skalenwert

1) SIXTL, F.: Meßmethoden der Psychologie. Theoretische Grundlagen und Probleme, Weinheim 1967, S. 44

2) EDWARDS, A.L.: Techniques of Attitude Scale Construction, New York 1957, S. 20

3) Die weitere Darstellung hält sich an: EDWARDS, A.L., a.a.O., S. 20 ff. und TORGERSON, W.S.: Theory and Methods of Scaling, New York/London/Sydney 1967⁷⁾, S. 159 ff.

4) SIXTL, F., a.a.O., S. 385 ff.

auf dem psychologischen Kontinuum zuzuordnen.

Konfrontiert man eine Versuchsperson mit dem Stimulus i , so löst dieser bei der Vp einen modalen Diskriminationsprozeß (modal discriminial process) auf dem psychischen Kontinuum aus. Ein mehrmaliges 'Vorlegen' des Stimulus zur Beurteilung durch die Vp wird zwar nicht jedes Mal denselben Diskriminationsprozeß auslösen, doch die Verteilung aller Diskriminationsprozesse wird der GAUSSschen Normalverteilungskurve entsprechen. Bekanntlich haben Mittelwert, Median und Modalwert in der Normalverteilung denselben Zahlenwert. Der Mittelwert entspricht dem Skalenwert i ($\bar{S}_i$). Die Streuung der Diskriminationsprozesse (discriminal dispersion) entspricht der Standardabweichung des Stimulus i . Die Streuungen verschiedener Stimuli sind in der Regel nicht gleich groß.

Beim Paarvergleich werden dem Beurteiler jeweils zwei Behauptungen, i und j , mit der Aufforderung präsentiert, jeweils zu entscheiden, welches item eine positivere Einstellung z.B. gegenüber den Engländern, dem 'psychologischen Objekt', repräsentiere. Da nun jedes item einer genügend großen Zahl von Experten zur Beurteilung vorgelegt wird, erhält man eine Verteilung der Urteile, die eine Entscheidung erlauben, ob das item i eine positivere Ladung habe als das item j . Ziehen z.B. 50% der Experten das item i dem item j vor ($i > j$), der Rest der Beurteiler aber das item j dem item i ($j > i$), dann können wir den Schluß ziehen, daß die durch i und j ausgelösten Diskriminationsprozesse gleich sind, also $\bar{S}_i = \bar{S}_j$. Erhalten wir weniger als 50% der Expertenurteile für i , dann ist auch der Skalenwert von i geringer: $\bar{S}_i < \bar{S}_j$, "The scale separation of the modal discriminial processes, $\bar{S}_i$ und $\bar{S}_j$, on the psychological continuum should, in other words, be some function of the proportion of judgments i greater than j ."¹⁾

Bei $n=14$ Behauptungen muß jeder Experte $n(n-1) / 2 = 91$ Urteile abgeben. Jedes item-Paar wird einer mittleren Anzahl ($n > 30$) von Beurteilern vorgelegt. Es folgt die Auszählung der Häufigkeit, mit der item i positiver bewertet wurde als item j : $f_{ij} = i > j$ (1)

1) EDWARDS, A.L., a.a.O., S. 24

Dividieren wir die Häufigkeit f_{ij} durch die totale Anzahl der Urteile, so erhalten wir den Prozentsatz

$$p_{ij} = \frac{f_{ij}}{N} \quad (2),$$

wobei p_{ij} normiert werden kann als z_{ij} .

Wir hatten angenommen, daß die durch die Stimuli i und j ausgelösten Diskriminationsprozesse sich normal verteilen. Man kann nun zeigen, daß die Differenzen zwischen zwei normal verteilten Variablen sich ebenfalls normal verteilen. Die Standardabweichung der Differenzen der Stimuli i und j kann nach der Formel

$$s_{i-j} = \sqrt{s_i^2 + s_j^2 - 2 \cdot r_{ij} \cdot s_i \cdot s_j} \quad (3)$$

berechnet werden. So erhalten wir die Standardwerte

$$z_{ij} = \frac{\bar{s}_i - \bar{s}_j}{s_{i-j}}, \text{ d.h., der Standardabstand} \quad (4)$$

zwischen den zwei modalen Diskriminationsprozessen $\bar{s}_i$ und $\bar{s}_j$ beträgt dann

$$\bar{s}_i - \bar{s}_j = z_{ij} \sqrt{s_i^2 + s_j^2 - 2 \cdot r_{ij} \cdot s_i \cdot s_j} \quad (5)$$

$$z_{ij} = \frac{\bar{s}_i - \bar{s}_j}{\sqrt{s_i^2 + s_j^2 - 2 \cdot r_{ij} \cdot s_i \cdot s_j}} \quad (6)$$

Für die drei Unbekannten s_i , s_j und r_{ij} in der Gleichung (6) kann "man in der Regel keine empirischen Werte gewinnen"¹⁾. Um die Gleichung lösbar zu machen, werden einige Annahmen gemacht: (1) $r_{ij} = 0$, woraus $\bar{s}_i - \bar{s}_j = z_{ij} \sqrt{s_i^2 + s_j^2}$ folgt, (2) $s_i = s_j = s$, d.h. die Annahme gleich großer Streuungen der Reaktionsverteilungen vereinfacht die Gleichung (5) zu $\bar{s}_i - \bar{s}_j = z_{ij} \cdot s \sqrt{2}$. Die multiple Konstante $s \sqrt{2}$ kann vernachlässigt werden, so daß die Standardwerte z_{ij} als Abstände zwischen den Stimuli interpretiert werden können.

1) vgl. den mit dem Vektorkalkül geführten Beweis bei GEER, J.P. van de: Introduction to Multivariate Analysis for the Social Sciences, San Francisco 1971, S. 19 ff.
2) SIXTL, F., a.a.O., S. 52

4.5.1.3. Konstruktion der Einstellungsskala 'Englisch'

In den Lehrbüchern werden eine Reihe von Skalierungsmethoden beschrieben; doch die wichtige Frage, "welches Verfahren man für eine gegebene Forschungsfrage wählen soll"¹⁾, wird kaum beantwortet.

Am besten geeignet erschien uns die Analyse latenter Strukturen von P.F.LAZARSFELD, da diese Methode sich durch den "explizite(n) Rückbezug(s) des Modells auf die Konzeption der Dispositionstermini bzw. intervenierenden Variablen"²⁾ auszeichnet, doch setzt dieses Verfahren mathematische Spezialkenntnisse und den Zugang zu einer Großrechenanlage voraus.³⁾ Diese beiden Bedingungen konnten wir nicht erfüllen.

Für die Auswahl des Paarvergleichs sprachen fünf Argumente:

1. Der Paarvergleich ist zuverlässig und bietet die Möglichkeit, typische Skalierungsfehler zu vermeiden.
2. "Obwohl nun beim Paarvergleich für die Rangordnungen keine Maßstab vorgegeben wird, liegt doch den schließlichen Anordnungen eine unabhängiger, durchgehender Maßstab zugrunde - indem nämlich nach der Versuchsanordnung jeder Fall Maßstab für alle anderen ist, wobei sich durch die Rotation der Fälle ein Kontinuum ergibt, als ob ein Fall in eine bereits geordnete Folge von Stimuli einzufügen wäre."⁴⁾ Der Stimulus, welcher jeweils mit den (n-1) anderen Stimuli verglichen wird, dient dabei als Anker.⁵⁾
3. Der Paarvergleich erlaubt eine fundamentale Messung.⁶⁾
4. Der Paarvergleich führt mit großer Wahrscheinlichkeit zu einer Intervallskala.⁷⁾

-
- 1) SCHEUCH, E.K.: Skalierungsverfahren in der Sozialforschung, in: König, R.(Hrsg.): Handbuch der empirischen Sozialforschung, Band I, Stuttgart 1967², S. 376
 - 2) SCHEUCH, E.K., a.a.O., S. 366
 - 3) SCHEUCH, E.K., a.a.O., S. 365
 - 4) SCHEUCH, E.K., a.a.O., S. 355 - 356
 - 5) Zum Begriff des Ankers vgl. SIXTL, F.(1967), a.a.O., S. 61 ff.
 - 6) SUPPES, P./ ZINNES, J.L.: Basic measurement theory, in: Luce, R.D./Bush, R.R./Galanter, E. (eds.): Handbook of mathematical psychology, Vol. I, New York/London 1963, S. 3 - 76
 - FISCHER, G.(Hrsg.): Psychologische Testtheorie, Bern/Stuttgart 1968
 - FRICKE, R.: Über Meßmodelle in der Schulleistungsdiagnostik, Düsseldorf 1972
 - 7) vgl. dazu die Einschränkung bei ESSER, U.: Skalierungsverfahren, in: Friedrich, W. (Hrsg.): Methoden der marxistisch-leninistischen Sozialforschung, Berlin 1971

5. "Das Verfahren des Paarvergleichs ist heute eines der am meisten mathematisch und methodisch ausgearbeiteten Verfahren der Merkmalsskalierung."¹⁾

Aus den schon oben genannten $n=14$ Behauptungen (Indikatoren) sollte eine Einstellungsskala konstruiert werden. Nach der Kombinatorik können $n(n-1)/2=91$ Paare von Behauptungen den Experten vorgelegt werden. Als Experten dienten 50²⁾ Studenten der PH Bielefeld.

Bei dem Vergleich zweier Behauptungen werden Ordinalurteile abgegeben: $a > b$ oder $b < a$. Urteile der Form $a=b$ sind nicht zugelassen. Stimmen alle Beurteiler oder keiner einer Aussage zu, so erhält man für die Aussage Prozentwerte von 100 oder 0. Wie aber aus der Gleichung (5) zu ersehen ist, werden die Stimuliabstände aus den Standardwerten z_{ij} gewonnen. Prozentwerten von 100 oder 0 entsprechen aber z -Werte, die von "minus Unendlich bis plus Unendlich reichen, was zu erheblichen Fehlern in der Berechnung der Skalenwerte führt."³⁾ Um die Skalenwerte, die aus solchen unvollständigen Datenmatrizen stammen, berechnen zu können, hat man ein statistisches Kalkül⁴⁾ zur Lösung dieses Problems entwickelt.

Nach dem bei HOFSTÄTTER beschriebenen Verfahren werden die Spalten der Prozentmatrix so geordnet, daß die Spaltensummen eine aufsteigende Rangfolge bilden. In der dazu gehörenden Z -Matrix bleiben die Zellen leer, die plus oder minus Unendlich enthalten. EDWARDS⁵⁾ gibt als Kriterium für ein Leerlassen der Zellen in der Z -Matrix einen Prozentwert von 98% oder weniger als 2% an, falls weniger als 200 Beurteiler vorhanden waren.

Die folgenden Tabellen geben die Häufigkeiten (F -Matrix), die daraus errechneten Prozentwerte (P -Matrix) und die den Prozentwerten korrespondierenden z -Werte (Z -Matrix) wieder.

1) ESSER, U., a.a.O., S. 201

2) EDWARDS, A.L., a.a.O., S. 28, fordert 50 - 100 Experten; Instruktion für die Studenten im Anhang

3) ESSER, U., a.a.O., S. 204

4) HOFSTÄTTER, P.R./WENDT, D.: Quantitative Methoden der Psychologie. Eine Einführung, München 1967³⁾, S. 243 ff.

5) EDWARDS, A.L., a.a.O., S. 41; vgl. GUILFORD, J.P.: Psychometric Methods, New York 1954²⁾, S. 163

Tabelle 6: F-Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	25	41	31	47	13	14	50	49	35	2	7	12	48	42
2	9	25	2	44	6	3	48	47	18	1	2	6	48	42
3	19	48	25	47	4	8	46	47	21	0	4	7	47	41
4	3	6	3	25	1	4	34	30	4	0	0	2	31	25
5	37	44	46	49	25	37	49	49	44	2	25	33	48	44
6	36	47	42	46	13	25	47	47	42	0	9	21	47	44
7	0	2	4	16	1	3	25	30	5	2	3	3	28	23
8	1	3	3	20	1	3	20	25	3	1	3	3	31	29
9	15	32	29	46	6	8	45	47	25	2	2	5	46	41
10	48	49	50	50	48	50	48	50	48	25	47	46	48	47
11	43	48	46	50	25	41	47	49	48	3	25	47	45	23
12	38	44	43	48	17	29	47	47	45	4	3	25	44	47
13	2	2	3	19	2	3	22	19	4	2	5	6	25	48
14	8	8	9	25	6	6	27	21	9	3	27	3	2	25

Tabelle 7: P-Matrix (geordnet)

	10	11	5	12	6	1	3	9	2	14	4	13	7	8
10	50	94	96	92	100	96	100	96	98	94	100	96	96	100
11	6	50	50	94	82	86	92	96	96	46	100	90	94	98
5	4	50	50	66	74	74	92	88	88	88	98	96	98	98
12	8	6	34	50	58	76	86	90	88	94	96	88	94	94
6	0	18	26	42	50	72	84	84	94	88	92	94	94	94
1	4	14	26	24	28	50	62	70	82	84	94	96	100	98
3	0	8	8	14	16	38	50	42	96	82	94	94	92	94
9	4	4	12	10	16	30	58	50	64	82	92	92	90	94
2	2	4	12	12	6	18	4	36	50	84	88	96	96	94
14	6	54	12	6	12	16	18	18	16	50	50	4	54	42
4	0	0	2	4	8	6	6	8	12	50	50	62	68	60
13	4	10	4	12	6	4	6	8	4	96	38	50	44	38
7	4	6	2	6	6	0	8	10	4	46	32	56	50	60
8	0	2	2	6	6	2	6	6	6	58	40	62	40	50

92 320 336 438 468 568 672 702 798 1042 1064 1076 1110 1114

Tabelle 8: Z-Matrix (geordnet)

	10	11	5	12	6	1	3	9	2	14	4	13	7	8
10	0.00	1.56	1.75	1.41	-	1.75	-	1.75	2.05	1.56	-	1.75	1.75	-
11	-1.56	0.00	0.00	1.56	0.92	1.08	1.41	1.75	1.75	-0.10	-	1.28	1.56	2.05
5	-1.75	0.00	0.00	0.41	0.64	0.64	1.41	1.18	1.18	1.18	2.05	1.75	2.05	2.05
12	-1.41	-1.56	-0.41	0.00	0.20	0.71	1.08	1.28	1.18	1.56	1.75	1.18	1.56	1.56
6	-	-0.92	-0.64	-0.20	0.00	0.58	0.99	0.99	1.56	1.18	1.41	1.56	1.56	1.56
1	-1.75	-1.08	-0.64	-0.71	-0.58	0.00	0.31	0.52	0.92	0.99	1.56	1.75	-	2.05
3	-	-1.41	-1.41	-1.08	-0.99	-0.31	0.00	-0.20	1.75	0.92	1.56	1.56	1.41	1.56
9	-1.75	-1.75	-1.18	-1.28	-0.99	-0.52	0.20	0.00	0.36	0.92	1.41	1.41	1.28	1.56
2	-2.05	-1.75	-1.18	-1.18	-1.56	-0.92	-1.75	-0.36	0.00	0.99	1.18	1.75	1.75	1.56
14	-1.56	0.10	-1.18	-1.56	-1.18	-0.99	-0.92	-0.92	0.99	0.00	0.00	1.75	0.10	-0.20
4	-	-	-2.05	-1.75	-1.41	-1.56	-1.56	-1.41	-1.18	0.00	0.00	0.31	0.47	0.25
13	-1.75	-1.28	-1.75	-1.18	-1.56	-1.75	-1.56	-1.41	-1.75	1.75	-0.31	0.00	-0.15	-0.31
7	-1.75	-1.56	-2.05	-1.56	-1.56	-	-1.41	-1.28	-1.75	-0.10	-0.47	0.15	0.00	0.25
8	-	-2.05	-2.05	-1.56	-1.56	-2.05	-1.56	-1.56	-1.56	0.20	-0.25	0.31	-0.25	0.00

Um die Skalenwerte der Stimuli 1 bis 14 zu erhalten, müssen folgende Rechenschritte unter-
nommen werden:

1. Bildung der Differenzen zweier nebeneinanderliegender Zahlen aus der Z-Matrix, d.h. in jeder Spalte wird von jeder Zahl die linke Nachbarzahl abgezogen: $1.56 - (+0.00) = 1.56$.
2. Das arithmetische Mittel einer jeden Spalte wird errechnet.
3. Der Nullpunkt der Skala wird willkürlich festgelegt.

Tabelle 2: z - Differenzen - Matrix

	11-10	5-11	12-5	6-12	1-6	3-1	9-3	2-9	14-2	4-14	13-4	7-13	8-7
10	+1.56	+0.19	-0.34	-	-	-	-	+0.30	-0.49	-	-	0.00	-
11	+1.56	0.00	+1.56	-0.64	+0.16	+0.33	+0.34	0.00	-1.85	-	-	+0.28	+0.49
5	+1.75	0.00	+0.41	+0.23	0.00	+0.77	-0.23	0.00	0.00	+0.87	-0.30	+0.30	0.00
12	-0.15	+1.15	+0.41	+0.20	+0.51	+0.37	+0.20	-0.10	+0.38	+0.19	-0.57	+0.38	0.00
6	-	+0.28	+0.44	+0.20	+0.58	+0.41	0.00	+0.57	-0.38	+0.23	+0.15	0.00	0.00
1	+0.67	+0.44	-0.07	+0.13	+0.58	+0.31	+0.21	+0.40	+0.07	+0.57	+0.19	-	-
3	-	0.00	+0.33	+0.09	+0.68	+0.31	-0.20	+1.95	-0.83	+0.64	0.00	-0.15	+0.15
9	0.00	+0.57	-0.10	+0.29	+0.47	+0.72	-0.20	+0.36	+0.56	+0.49	0.00	-0.13	+0.28
2	+0.30	+0.57	0.00	-0.38	+0.64	-0.83	+1.39	+0.36	+0.99	+0.19	+0.57	0.00	-0.19
14	+1.66	-1.28	-0.38	+0.38	+0.19	+0.07	0.00	-0.07	+0.99	0.00	-1.75	+1.85	-0.30
4	-	-	+0.30	+0.34	-0.15	0.00	+0.15	+0.23	+1.18	0.00	+0.31	+0.16	-0.22
13	+0.47	-0.47	+0.57	-0.38	-0.19	+0.19	+0.15	-0.34	0.00	-2.06	+0.31	-0.15	-0.16
7	+0.19	-0.49	+0.49	0.00	-	-	+0.13	-0.47	+1.65	-0.37	+0.62	-0.15	+0.25
8	-	0.00	+0.49	0.00	-0.49	+0.49	0.00	0.00	+1.76	-0.45	+0.56	-0.56	+0.25
Sa	8.01	0.96	4.11	0.46	2.98	3.14	1.94	3.19	4.03	0.30	0.09	1.83	0.55
$\bar{x}$	0.801	0.073	0.293	0.035	0.248	0.261	0.149	0.227	0.287	0.025	0.007	0.140	0.045
$\bar{s}_{10} = \bar{s}_{11} = \bar{s}_5 = \bar{s}_{12} = \bar{s}_6 = \bar{s}_1 = \bar{s}_3 = \bar{s}_9 = \bar{s}_2 = \bar{s}_{14} = \bar{s}_4 = \bar{s}_{13} = \bar{s}_7 = \bar{s}_8 =$	1.000	1.801	1.874	2.202	2.450	2.711	2.860	3.087	3.374	3.399	3.406	3.546	3.591

Vor der indikator-orientierten Skalierung mit Studenten als Experten sammelten wir die Daten für eine reaktions-orientierte Skalierung an Hauptschulen in Belecke, Körbecke und Soest (153 Schüler; 5. Schuljahr), Regierungsbezirk Arnsberg. Wir hielten dieses Vorgehen für methodisch sinnvoller bei Schülern dieser Altersstufe, die sich kaum von ihrer eigenen Einstellung¹⁾ distanzieren können. Methodisch berücksichtigt wurde dies in der Instruktion:

"In diesem Jahr beginnt für Euch der Englischunterricht. Mir geht es darum zu erfahren, was Ihr von England, den Engländern und der englischen Sprache haltet. Deshalb findet Ihr auf den folgenden Seiten eine Anzahl von Behauptungen zu dem genannten Bereich.

Jeweils zwei Behauptungen sollt Ihr miteinander vergleichen. Am Rande kreuzt Ihr dann die Behauptung an, der Ihr am meisten zustimmt.

Stimmenthaltungen sind nicht zugelassen."

Tabelle 10: F-Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	-	118	110	84	67	94	136	117	124	15	59	80	117	119
2	35	-	49	79	44	71	122	101	114	15	51	55	122	121
3	43	104	-	59	57	73	131	107	114	17	48	52	121	113
4	69	74	94	-	50	80	136	112	115	26	61	62	128	122
5	86	109	96	103	-	75	127	120	123	27	77	82	122	118
6	59	82	80	73	78	-	119	120	127	28	65	80	130	121
7	17	31	22	17	26	34	-	63	73	17	35	38	85	94
8	36	52	46	41	33	33	90	-	66	23	38	42	95	97
9	29	39	39	38	30	26	80	87	-	21	39	52	106	115
10	138	138	136	127	126	125	136	130	132	-	123	124	131	135
11	94	102	105	92	76	88	118	115	114	30	-	84	121	121
12	73	98	101	91	71	73	115	111	101	29	69	-	130	128
13	36	31	32	25	31	23	68	58	47	22	32	23	-	82
14	34	32	40	31	35	32	59	56	38	18	32	25	71	-

1) Zum Problem, ob die individuelle Position die Skalierung beeinflusst, vgl. OERTER, R.: Struktur und Wandlung von Werthandlungen, München 1970, S. 94 ff. und SIXTL, F., a.a.O., S. 250 ff.

Tabelle 11: P-Matrix (geordnet)

	10	5	11	1	12	6	4	3	2	9	8	7	13	14
10	-	82	80	90	81	82	83	89	90	86	85	89	86	88
5	18	-	50	56	54	49	67	63	71	80	78	83	80	77
11	20	50	-	61	55	58	60	69	67	75	75	77	79	79
1	10	44	39	-	52	61	55	72	77	81	76	89	76	78
12	19	46	45	48	-	48	59	66	64	66	73	75	85	84
6	18	51	42	39	52	-	48	52	54	83	78	78	85	79
4	17	33	40	45	41	52	-	61	48	75	73	89	84	80
3	11	37	31	28	34	48	39	-	68	75	70	86	79	74
2	10	29	33	23	36	46	52	32	-	75	66	80	80	79
9	14	20	25	19	34	17	25	25	25	-	57	52	69	75
8	15	22	25	24	27	22	27	30	34	43	-	59	62	63
7	11	17	23	11	25	22	11	14	20	48	41	-	56	61
13	14	20	21	24	15	15	16	21	20	31	38	44	-	54
14	12	23	21	22	16	21	20	26	21	25	37	39	46	-
%	189		475		522		562		659		847		967	
%		474		490		541		615		843		940		971

Die Skalenwerte der 14 Stimuli wurden bei der Expertengruppe (Studenten) aus einer unvollständigen Datenmatrix geschätzt, während bei der Schülerstichprobe eine vollständige Datenmatrix vorlag. Nach der zuerst genannten Methode erhält man "the scale separations of adjacent stimuli", während bei einer Berechnung der Skalenwerte aus einer vollständigen Matrix diese ausgedrückt werden "in terms of their deviations from the mean of all of the scale values."¹⁾

1) EDWARDS, A.L., a.a.O., S. 46

Tabelle 12: Z-Matrix (geordnet)

	10	5	11	1	12	6	4	3	2	9	8	7	13	14
10	-	0.92	0.84	1.28	0.88	0.92	0.95	1.23	1.28	1.08	1.04	1.23	1.08	1.18
5	-0.92	-	0.00	0.15	0.10	-0.03	0.44	0.33	0.55	0.84	0.77	0.95	0.84	0.74
11	-0.84	0.00	-	0.28	0.13	0.20	0.25	0.50	0.44	0.67	0.67	0.74	0.81	0.81
1	-1.28	-0.15	-0.28	-	0.05	0.28	0.13	0.58	0.74	0.88	0.71	1.23	0.71	0.77
12	-0.88	-0.10	-0.13	-0.05	-	-0.05	0.23	0.41	0.36	0.41	0.61	0.67	1.04	0.99
6	-0.92	0.03	-0.20	-0.28	0.05	-	-0.05	0.05	0.10	0.95	0.77	0.77	1.04	0.81
4	-0.95	-0.44	-0.25	-0.13	-0.23	0.05	-	0.28	-0.05	0.67	0.61	1.23	0.99	0.84
3	-1.23	-0.33	-0.50	-0.58	-0.41	-0.05	-0.28	-	0.47	0.67	0.52	1.08	0.81	0.64
2	-1.28	-0.55	-0.44	-0.74	-0.36	-0.10	0.05	-0.47	-	0.67	0.41	0.84	0.84	0.81
9	-1.08	-0.84	-0.67	-0.88	-0.41	-0.95	-0.67	-0.67	-0.67	-	0.18	0.05	0.50	0.67
8	-1.04	-0.77	-0.67	-0.71	-0.61	-0.77	-0.61	-0.52	-0.41	-0.18	-	0.23	0.31	0.33
7	-1.23	-0.95	-0.74	-1.23	-0.67	-0.77	-1.23	-1.08	-0.84	-0.05	-0.23	-	0.15	0.28
13	-1.08	-0.84	-0.81	-0.71	-1.04	-1.04	-0.99	-0.81	-0.84	-0.50	-0.31	-0.15	-	0.13
14	-1.18	-0.74	-0.81	-0.77	-0.99	-0.81	-0.84	-0.64	-0.81	-0.67	-0.33	-0.28	-0.13	-
Sa	-13.91	-4.76	-4.46	-4.37	-3.51	-3.12	-2.62	-0.81	0.32	5.44	5.42	8.59	8.99	9.00
$\bar{X}$	-0.99	-0.34	-0.32	-0.31	-0.25	-0.22	-0.19	-0.06	0.02	0.39	0.39	0.61	0.64	0.64
$\bar{X}+K$	1.00	1.65	1.67	1.68	1.74	1.77	1.80	1.93	2.01	2.38	2.38	2.60	2.63	2.63

4.5.1.4. Evaluation der Einstellungsskala 'Englisch'

Die Evaluierung der Einstellungsskala 'Englisch' schließt alle die Maßnahmen ein, welche auf die Feststellung der Standardkriterien (1) Objektivität, (2) Reliabilität und (3) Validität gerichtet worden sind.

- (1) Durchführungsobjektivität wird durch die standardisierte Instruktion erreicht; die Auswertungsobjektivität ist trivial.
- (2) Um die Reliabilität des Meßinstruments zu überprüfen, führten wir im September 1973 an vier Hauptschulen in Lippstadt eine Reliabilitätsuntersuchung nach der Test-Retest-Methode durch. Diese Methode ist nicht unproblematisch. "Zu der häufig gestellten Frage, wieviel Zeit zwischen dem Test und seiner Wiederholung liegen soll, läßt sich nicht eindeutig Stellung nehmen. Man befindet sich hier in einem gewissen Dilemma: Einmal sollen die Erinnerungsspuren möglichst verblassen, das Intervall also möglichst groß sein; zum anderen soll das untersuchte Persönlichkeitsmerkmal unverändert bleiben, was wiederum ein kurzes Intervall wünschenswert erscheinen läßt. Je nach der Eigenart des Testinhaltes werden die Zeitabstände zwischen Test und Wiederholung von Tagen bis zu Monaten variieren. Gleichgültig, wie groß das Intervall ist, Schwankungen des Persönlichkeitsmerkmals werden in der Regel in den Wiederholungskoeffizienten eingehen. In diesem Sinne macht dieser Koeffizient auch eine Aussage über die zeitliche Stabilität des untersuchten Persönlichkeitsmerkmals."¹⁾

Die Berechnung des Stabilitätskoeffizienten erfolgte nach der Formel des Produkt-Moment-Korrelationskoeffizienten, dessen Voraussetzungen - Linearität, Normalität und Intervallskalenniveau der Variablen - durch die Daten erfüllt sind:

$$r_{xy} = \frac{\sum xy}{N \cdot s_x \cdot s_y} \quad 2)$$

-
- 1) LIENERT, G.A.: Testaufbau und Testanalyse, Weinheim/Berlin 1967², S. 217 - 218
 - 2) GUILFORD, J.P.: Fundamental statistics in psychology and education, New York/Sydney 1965⁴, S. 95

Wir erhielten folgende Koeffizienten:

Klasse A (n=28): r=0.72 Klasse C (n=19): r=0.91
Klasse B (n=26): r=0.74 Klasse D (n=25): r=0.69

Die geringe Anzahl von Schülern pro Klasse ist darauf zurückzuführen, daß erstens einige Schüler bei der Testwiederholung fehlten und zweitens die Gastarbeiterkinder nicht berücksichtigt worden sind.

- (3) Nimmt man die Expertenurteile als Kriteriumsvariable, so bietet sich eine Methode der Validitätsüberprüfung geradezu an, nämlich die Korrelation der Rangreihen der Stimuli der indikator- und reaktionsorientierten Messung.

Tabelle 13: Rangfolge der Behauptungen: (a) nach dem Schüler- und (b) nach dem Studentenurteil

item	1	2	3	4	5	6	7	8	9	10	11	12	13	14
(a)	11	6	7	8	13	9	3	4	5	14	12	10	2	1
(b)	9	6	8	4	12	10	2	1	7	14	13	11	3	5

Die Korrelation nach dem SPEARMANSchen Rangkorrelationskoeffizienten

$$r_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N^3 - N}$$

ergab die Übereinstimmung von $r_s = 0.877$. Eliminiert man die items 4 und 14, so erhält man die hohe Korrelation von $r_s = 0.948$. Die Einstellungsskala 'Englisch' ist also ein objektives, reliables und valides Meßinstrument.

4.5.2. Das Semantische Differential

Nicht auf das Endprodukt kognitiver Lernprozesse wie z.B. faktisches Wissen in einem Stoffgebiet richtet sich unser Forschungsinteresse, sondern die Nebenwirkungen normorientierter Prüfinstrumente im affektiven Bereich, wie sie auf die von uns für relevant gehaltenen Konzepte durchschlagen, standen im Mittelpunkt unserer Bemühungen. Das Semantische Differential eignet sich be-

1) SIEGEL, S.: Nonparametric Statistics, New York 1956, S. 204

sonders dafür, "emotionale Bedeutungsprozesse (indirekt) quantitativ zu erfassen."¹⁾

4.5.2.1. Methodische und methodologische Probleme

Der Vorteil der oben beschriebenen THURSTONEschen Skalierungsmethode besteht darin, daß die Experten die vom Versuchsleiter vorgegebenen Skalen nicht als Zwangsjacke bei der Einordnung von Reizen benutzen müssen, ein Nachteil ist jedoch darin zu sehen, daß die postulierte Eindimensionalität der Skala in den seltensten Fällen zu erreichen ist. Eine so konstruierte Skala ist "vermutlich nicht eindimensional (z.B. pro- und antisemitisch). Vielmehr erscheint es sinnvoll, jedes soziale Urteil von vornherein als mehrdimensional, d.h. aus verschiedenen Komponenten zusammengesetzt, aufzufassen. Der Gegenstand des Urteils selbst, nämlich der soziale Reiz (Personen, Behauptungen über soziale Beziehungen, Situationen) ist so komplex, daß man bei seiner Darbietung nicht eine Variation auf einer einzigen Dimension vornehmen kann. Vielmehr ist hierbei weitgehend unbekannt, was eigentlich variiert wird. Man hat bei der Beurteilung sozialer Reize also sowohl auf seiten des Urteils als auch auf der Reizseite selbst Multidimensionalität vor sich."²⁾

Die Multidimensionalität von Urteilen wird in dem Ansatz von OSGOOD, Suci und TANNENBAUM³⁾ berücksichtigt. So kann man sich z.B. das Konzept 'Lehrer' als Punkt oder Vektor in einem mehrdimensionalen Raum denken. Faktorenanalysen haben drei Bedeutungsdimensionen ergeben: evaluation (Bewertung), potency (Stärke) und activity (Aktivität). Dabei wird der größte Teil der Varianz - fast soviel wie die beiden anderen Faktoren besitzen - durch den Faktor Bewertung aufgeklärt.⁴⁾ Die emotionale Natur dieses Raumes hebt ERTEL⁵⁾ hervor.

1) OERTER, R.: Struktur und Wandel von Werthaltungen, München 1970, S. 96; vgl. ROTH, E. (1967), a.a.O., S. 34

2) FUCHS, A.: Das Eindrucksdifferential als Instrument emotionaler Bedeutungsprozesse, in: BERGLER, R. (Hrsg.): Das Eindrucksdifferential. Theorie und Technik, Bern/Stuttgart/Wien 1975, S. 99

3) OSGOOD, Ch.E. u.a.: The measurement of meaning, Illinois 1971⁸

4) OSGOOD, Ch.E. u.a., S. 31 ff.

5) ERTEL, S.: Die emotionale Natur des semantischen Raumes, in: Psychologische Forschung 1964

Der postulierte semantische Raum ist ebenso wie die lerntheoretische Annahme eines Vermittlungsprozesses ein hypothetisches Konstrukt.¹⁾ Dieser Konstruktcharakter des Vermittlungsprozesses wird deutlich in der Definition:

"A pattern of stimulation which is not the significate is a sign of that significate if it evokes in the organism a mediating process, this process (a) being some fractional part of the total behavior elicited by the significate and (b) producing responses which would not occur without the previous contiguity of non-significate and significate patterns of stimulation."²⁾

Hermeneutische Probleme ergeben sich bei der Konstruktion eines semantischen Differentials:

- (1) Welches Hypothesennetz (welche Theorie) bestimmt die Auswahl der relevanten Begriffe (Konzepte)?
- (2) Welche Relevanz haben die Adjektivskalen (a) für die Begriffe und (b) für die Dimension des semantischen Raumes?

"The crux of the method, of course, lies in selecting the sample of descriptive polar terms."³⁾ Für unsere Untersuchung übernahmen wir die Adjektivpaare und Skalen, welche HURRELMANN "zur Erfassung der Einstellungsstereotype und der Selbsteinstellungen der Schüler"⁴⁾ entwickelt hat. Folgende Begriffe (Konzepte) sollten damit erfaßt werden: SELBSTBILD/WUNSCHBILD/FREMBBILD des guten und schlechten Schülers/ENGLISCHLEHRER/KLASSENARBEIT/ZENSUR/KLASSEN-TEST.

HURRELMANN hat faktorenanalytisch nachgewiesen, daß die Skalen auf den von OSGOOD u.a. gefundenen Dimensionen Bewertung, Stärke und Aktivität laden. Diesem Vorteil steht jedoch für unsere Untersuchung der Nachteil entgegen, daß die Adjektivpaare der Skalen nicht aufgrund einer phänomenologischen Bedeutungsanalyse der einzelnen Konzepte ausgewählt worden sind. Wir sind uns bewußt, daß dies einer Reduktion des Bedeutungsgehaltes der Begriffe entspricht.

1) Zur Kritik vgl. CICOUREL, A. von, a.a.O., S. 259

2) OSGOOD, Ch.E. u.a., a.a.O., S. 7

3) OSGOOD, Ch.E. u.a., a.a.O., S. 20

4) HURRELMANN, K.: Unterrichtsorganisation und schulische Organisation. Eine empirische Untersuchung zur Rolle der 'Leistungs-differenzierung' im schulischen Selektionsprozeß, Weinheim/Berlin/Basel 1971, S. 255 ff.

Bei unseren Voruntersuchungen stellte sich heraus, daß Schüler des 5. und 6. Schuljahres nach den Bedeutungen der Adjektive 'rege' und 'gemächlich' fragten.¹⁾ Die Klärung der Begriffe wurde deshalb am Tag vor der Hauptuntersuchung vorgenommen. Zur gleichen Zeit wurden die Schüler mit der Instruktion des Semantischen Differentials vertraut gemacht.

KLASSENTEST

fröhlich	3---2---1---0---1---2---3	traurig
weich	3---2---1---0---1---2---3	hart
freundlich	3---2---1---0---1---2---3	unfreundlich
gemächlich	3---2---1---0---1---2---3	eilig
ehrllich	3---2---1---0---1---2---3	unehrllich
nachgiebig	3---2---1---0---1---2---3	stark
gut	3---2---1---0---1---2---3	schlecht
ruhig	3---2---1---0---1---2---3	lebhaft
klug	3---2---1---0---1---2---3	dumm
zart	3---2---1---0---1---2---3	kraftvoll
angenehm	3---2---1---0---1---2---3	unangenehm
still	3---2---1---0---1---2---3	rege

Die Anzahl der Gegensatzpaare je Dimension ist bei HURRELMANN nicht gleich. Das kann eine Fehlerquelle sein, denn: "Wichtiger als die absolute Zahl der verwendeten Gegensatzpaare ist, daß je Dimension gleich viele Gegensatzpaare aufgenommen werden."²⁾

Die Auswertung des Semantischen Differentials mit Hilfe von Signifikanztests oder deskriptiven statistischen Verfahren erfordert in der Regel einen Datensatz mit Intervallskalenniveau. Die Annahme äquidistanter Intervalle beim Semantischen Differential erscheint jedoch fragwürdig, so daß Verfahren, in deren Modellprämissen gleiche Skalenabstände vorausgesetzt werden, höchstens zu deskriptiven Zwecken herangezogen werden dürften.

1) Bei HURRELMANN findet man keinen Hinweis auf derartige Schwierigkeiten

2) MAYNTZ, R. u.a., a.a.O., S. 52

Dies gilt z.B. auch für das von OSGOOD u.a. angegebene

Distanzmaß: $D_{11} = \sqrt{\sum_j d_{11}^2}$ ¹⁾).

Auch gegenüber den Faktorenanalysen, die mit den Daten Semantischer Differentiale gemacht worden sind, gelten unsere Bedenken. Wegen der offensichtlichen Verletzung der Prämissen statistischer Verfahren gewinnt man vielleicht Merkmalsräume, die nur Artefakte darstellen. Für ein bestimmtes Problem hat dies ORLIK²⁾ nachgewiesen. Er schlägt deshalb eine Transformation der Daten vor. Zu dieser Art von Transformationen bemerkt LORGE lakonisch: "Linear transformations of raw scores $z = X - M/s$ will not make the raw score something it never was."³⁾

In der Bewertung des Semantischen Differentials schließen wir uns dem Urteil SCHEUCHs an: "Umstritten ist, was das Verfahren nun eigentlich mißt - wobei diese Kontroverse für die Messung von Einstellungen nur teilweise relevant ist. Mathematisch ist das Verfahren zweifellos noch nicht abgeklärt. Nicht bewiesen ist die Annahme des Verfahrens, daß es sich bei den als Alternativen ausgewählten Attribut-Paaren im psychologischen Sinne um echte Gegensätze handelt, und insbesondere, ob diese Gegensätze die gleichen Abstände vom Nullpunkt ihres Kontinuums besitzen. Ein wichtiger Vorteil ist der mehrdimensionale Charakter der Technik und die Leichtigkeit der Administration, wenngleich dies erkauft wird um den Preis einer (wenn voll ausgeführt) umfangreichen Analyse."⁴⁾

4.2.2.2. Das Rationale des Semantischen Differentials

Konstruktion, Administration und Auswertungsmethoden des Semantischen Differentials sind in den meisten Lehrbüchern der empirischen Sozialforschung⁵⁾ dargestellt, so daß sich hier eine

1) OSGOOD, Ch. u.a., a.a.O., S. 91

2) ORLIK, P.: Eine Technik zur erwartungstreuen Skalierung psychologischer Merkmalsräume aufgrund von Polaritätsprofilen, in: Z.f. experim. angew. Ps., 1967; vgl. ORLIK, P.: Kritische Untersuchungen zur Begabtenförderung, Düsseldorf 1965

3) LORGE, I.: The Fundamental Nature of Measurement, in: Lindquist, E.F.(ed.): Educational Measurement, Washington 1963, S.556

4) SCHEUCH, E.K., a.a.O., S. 370

5) KERLINGER, F.N.: Foundations of Behavioral Research, New York 1973², S. 566 ff.

ausführliche Wiedergabe dieser Technik erübrigt. Deshalb fassen wir kurz das Rationale des Semantischen Differentials zusammen:

- (1) Es wird ein semantischer Raum unbekannter Dimensionalität, der jedoch Euklidischen Charakter hat, postuliert.
- (2) Die Bedeutung eines Begriffes (Konzepts) wird durch einen Punkt im multidimensionalen Raum repräsentiert.
- (3) Jede Skala besteht aus einem polaren Adjektivpaar. Eine Stichprobe solcher Skalen repräsentiert einen multidimensionalen Raum, der nach Faktorenanalysen hauptsächlich drei orthogonale Dimensionen (evaluation, potency, activity) ergeben hat.

4.5.3. Fragebogen für 'Prüfungsangst' und 'allgemeine Angst'

Das Definitionsproblem, was Angst sei, ist in unserer Untersuchung vorentschieden durch die Wahl der von SARASON u.a. entwickelten Fragebogen. "Das Studium der menschlichen Angst war von Anfang an durch das Fehlen einer allgemein anerkannten operationalen Definition des Konstrukts gehandikapt. Durch die Vielfalt der Definitionen ist es zu manchen Begriffsverwirrungen und zu Schwierigkeiten in der Interpretation widerspruchsvoller Versuchsergebnisse gekommen. Wenn die Ergebnisse zweier Untersuchungen, die sich nur durch die Definition des Konstrukts unterschieden, sich gegenseitig widersprechen, besteht immer die Möglichkeit, daß die Widersprüche von den verschiedenen Definitionen herrühren."¹⁾

Dem theoretischen Ansatz SARASONS liegen tiefenpsychologische Annahmen über die Genese von Angst zugrunde. SARASON übernimmt "die These von der Angst als Gefahrensignal, wie sie bei F r e u d , M o w r e e r u.a. konzipiert wurde, und versucht die psychoanalytischen Erkenntnisse F r e u dscher Prägung mit der H u l lschen drive-Theorie zu verbinden"²⁾, d.h. die Angsteinstellung wird schon im frühkindlichen Sozialisationsprozeß gelernt. Ursache des ängstlichen Verhaltens ist eine gestörte Eltern-Kind-Beziehung. Ontogenetisch früher erworbene Erfahrungen³⁾ - gespeichert als

1) LEVITT, E.E.: Die Psychologie der Angst, Stuttgart/Berlin/Mainz/Köln 1971, S. 140

2) SCHELL, H.: Angst und Schulleistung, Göttingen 1972, S.18

3) MÜHLE, G.: Furcht und Angst unter entwicklungspsychologischen Aspekt, in: Wiesbrock, H.(Hrsg.): Die politische und gesellschaftliche Rolle der Angst, 1967

unbewußte Inhalte¹⁾ - "werden in der Schulwirklichkeit bei Konstellationen, die der früheren Situation ähnlich sind, reaktualisiert"²⁾).

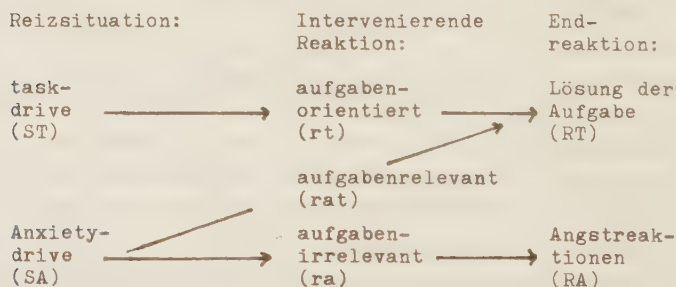
Die beiden das theoretische Konzept fundierenden Hypothesen SARASONS seien angeführt:

- "1. Die Reaktion des prüfungsängstlichen Kindes auf Prüfungen und prüfungsähnliche Situationen in der Klasse spiegelt seine Erfahrungen in psychologisch oder interpersonal ähnlichen Situationen in seinem Zuhause wider, vor und nach dem Beginn der formellen Einschulung.
2. Die Prüfungsangstreaktion hat, zusätzlich zu ihrer bewußten Bedeutung, eine gleichzeitige unbewußte Bedeutung, die sich vornehmlich auf die Erfahrungen bezieht, die das Kind in der familiären Situation gemacht hat."³⁾

Die Affinität der situationalen Komponenten von Schul- und häuslicher Erziehungssituation wird vom Kinde unbewußt erlebt. Die Mobilisierung von Affekten, ausgelöst in Prüfungssituationen, verhindert aufgabenrelevantes Verhalten. Aufgabemirrelevantes, affektiv bestimmtes Verhalten, das in der frühen Kindheit erworben worden ist, stellt - lerntheoretisch gesehen - in der Übertragung auf die schulische Prüfungssituation eine Generalisierung dar.

Das theoretische Modell der habit-Interferenz-Theorie von MANDLER und SARASON sei in der Abbildung 7 veranschaulicht.⁴⁾

Abb. 7: Habit-Interferenz-Theorie der Prüfungsangst von G. MANDLER und S.B. SARASON



- 1) Zur Problematik des Begriffs 'unbewußt' vgl. HOFSTÄTTER, P.R.: Tiefenpsychologische Persönlichkeits-Theorie, in: Thomae, H./Lersch, Ph.: Hdb. d. Ps., Bd. 4, Persönlichkeitsforschung und Persönlichkeitstheorie, Göttingen 1960², S. 594 ff.
- 2) SCHELL, H., a.a.O., S. 19
- 3) SARASON, S.B. u.a.: Angst bei Schulkindern. Ein Forschungsbericht, Stuttgart 1971, S. 22
- 4) Wir halten uns an die Skizze von SCHELL, H., a.a.O., S. 12

Ob dem mit der habit-Interferenz-Theorie von MANDLER und SARASON rivalisierenden drive-Modell von SPENCE und TAYLOR "eine höhere Validität zukommt, ist beim derzeitigen Stand der wissenschaftlichen Diskussion unmöglich"¹⁾ zu entscheiden.

4.5.3.1. TASC- Itemanalyse

Der Voruntersuchung von NICKEL, H. und SCHLÜTER, P. im Jahre 1970²⁾ folgte 1973 eine Nachuntersuchung an insgesamt 842 Haupt-, Real- und Oberschülern³⁾, die im wesentlichen mit den Ergebnissen der ersten Untersuchung übereinstimmte.

Die Prüfungsangstskala TASC enthält 29 items.⁴⁾ NICKEL, SCHLÜTER und FENNER paßten die Prüfungsangstskala TASC deutschen Verhältnissen an und führten eine Aufgabenanalyse durch, die wir tabellarisch zusammenfassen.

Tabelle 14: Koeffizienten der Aufgabenanalyse der TASC nach NICKEL, SCHLÜTER und FENNER (1973)

	Streubereich	Mittelwert
Schwierigkeitsgrad	$p_i = 0.17 - 0.78$	$\bar{p}_i = 0.44$
Trennschärfekoeffizient	$r_{it} = 0.29 - 0.64$	$\bar{r}_{it} = 0.51$

Die interne Konsistenz, ein Maß für die Homogenität der Skala, wurde nach der KUDER-RICHARDSON-FORMEL (8) errechnet⁵⁾ und ergab einen Koeffizienten der Größe $r_{tt} = 0.94$. Ein nach der split-half-Methode errechneter Koeffizient betrug $r_{tt} = 0.91$. Die Wiederholungszuverlässigkeit ergab im Durchschnitt $r_{tt} = 0.76$. Ihr Streuungsbereich erstreckt sich in den 18 Untergruppen von 0.55 bis 0.91. Die TASC erfüllt also die Anforderungen, die an ein standar-

1) SCHELL, H., a.a.O., S. 14

2) NICKEL, H./SCHLÜTER, P.: Angstwerte bei Hauptschülern und ihr Zusammenhang mit Leistungs- sowie Verhaltensmerkmalen, Lehrerurteil und Unterrichtsstil, in: Zeitschr. f. Entw.- und Pädag. Psychol., 1970

3) NICKEL, H./SCHLÜTER, P./FENNER, H.-J.: Angstwerte, Intelligenztest- und Schulleistungen sowie der Einfluß der Lehrerpersönlichkeit bei Schülern verschiedener Schularten, in: Psychologie in Erziehung und Unterricht, 1973

4) s. Anhang

5) LIENERT, G.A.: Testaufbau und Testanalyse, Weinheim/Berlin 1967²⁾, S. 226; HORST, P.: Messung und Voraussage. Eine Einführung in die psychologische Testtheorie, Weinheim/Berlin/Basel 1971, S. 324 ff.

disiertes Meßinstrument zu stellen sind.

4.5.3.2. CMAS - Itemanalyse

Die allgemeine Angstskala CMAS enthält 41 items.¹⁾ Die Itemanalyse brachte auch für die CMAS befriedigende Trennschärfe- und Schwierigkeitskoeffizienten, wie die Tabelle 15 zeigt:

Tabelle 15: Koeffizienten der Aufgabenanalyse der CMAS nach SCHLÜTER und FENNER (1973)

	Streubereich	Mittelwert
Schwierigkeitsgrad	$p_i = 0.15 - 0.83$	$\bar{p}_i = 0.48$
Trennschärfekoeffizient	$r_{it} = 0.29 - 0.55$	$\bar{r}_{it} = 0.40$

"Die interne Konsistenz liegt bei $r_{tt} = .92$ (KR 8), ferner wurde ein split-half-Koeffizient von $r_{tt} = .86$ berechnet (S - B aufgewertet), die Wiederholungszuverlässigkeit nach einem Monat beträgt im Durchschnitt $r_{tt} = .74$ und streut in den 18. Untergruppen von .54 bis .89."²⁾

Die Grenzen der Leistungsfähigkeit beider Skalen werden sichtbar, wenn man sich vor Augen hält, daß die Reaktionen der Vpn verbalem und nicht overttem Verhalten zuzuordnen sind.

4.5.3.3. Die Faktorenstruktur von CMAS und TASC

Um die von NICKEL u.a. (1973) 'gefundene' Faktorenstruktur aller items der CMAS und TASC zu relativieren, ist es methodologisch wichtig, sich der Annahmen zu vergewissern, die in den faktorenanalytischen Ansatz eingehen.

Das Fundamentaltheorem der Faktorenanalyse³⁾ für unkorrelierte Faktoren besagt, daß die Korrelationsmatrix R reproduziert werden kann, falls das Faktorenmuster A bekannt ist: $R = A A'$. Dann gilt für jedes einzelne Element von R

$$r_{ik} = a_{i1}a_{k1} + a_{i2}a_{k2} + \dots + a_{ir}a_{kr}$$

1) CASTANEDA, A/McCANDLESS, B.R./PALERMO, D.S.: The children's form of manifest anxiety scale, in: Child Developm., 1956, S. 317 - 326

2) NICKEL, H. u.a., a.a.O., S. 3

3) ÜBERLA, K.: Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler, Berlin/Heidelberg/New York 1971², S. 50 ff. und FRUCHTER, B.: Introduction to Factor Analysis, New Jersey 19688, S. 44 ff.

In diese Gleichung gehen drei Annahmen ein: (1) Linearität, d.h. ein Meßwert ist eine lineare Funktion anderer Größen, (2) Additivität und (3) Unabhängigkeit der additiven Größen.

Begnügt man sich mit der deskriptiven Funktion der Faktorenanalyse, der übersichtlichen Zusammenfassung umfangreicher Datensätze, so ist dies völlig unproblematisch, hypostasiert man jedoch die dem Kalkül entsprungenen Faktoren und schreibt ihnen die Eigenschaft der Kausalität zu, dann begibt man sich in den Bereich des Unbeweisbaren. Die Faktorenanalyse "analysiert lediglich die formalen Relationen der beobachteten Variablen und ist damit ganz auf die V o r e n t s c h e i d u n g (von mir hervorgehoben - K.H.) über die zu untersuchenden Variablen angewiesen. Eine Generalisierung der Faktoren ist immer nur auf die Grundgesamtheit von Variablen möglich, deren Stichprobe man der Analyse unterzogen hat."¹⁾

NICKEL u.a. erhielten nach einer Variomaxrotation vier Faktoren:

"a) J u n g e n :

Faktor I: A n g s t (i n d e r S c h u l e z u v e r -
s a g e n

TASC: Item Nr. 49, 53, 59, 46, 55, 43, 69, 64, 52,
50, 54, 51, 56

CMAS: Item Nr. 27, 3

Faktor II: A l l g e m e i n e , u n b e s t i m m t e A n g s t

CMAS: Item Nr. 35, 34, 41, 39, 33, 37, 36, 40, 48, 32

Faktor III: G e f ü h l (A n g s t v o r) s o z i a l e r
I s o l i e r u n g u n d E r w a r t u n g s -
A n g s t

CMAS: Item Nr. 10, 29, 9, 23, 16, 19, 25, 5, 13, 20

Faktor IV: A n g s t v o r s c h u l i s c h e r L e i -
s t u n g s p r ü f u n g

TASC: Item Nr. 67, 61, 70, 60, 66, 65, 71, 72, 69

b) M ä d c h e n :

Faktor I: A n g s t (i n d e r S c h u l e) z u v e r -
s a g e n

1) Überla, K., a.a.O., S. 255

TASC: Item Nr. 49, 59, 64, 55, 54, 46, 53, 56,
43, 52

CMAS: Item Nr. 32, 41, 34

Faktor II: Gefühl (Angst vor) sozialer
Isolierung und Erwartungs-
Angst

CMAS: Item Nr. 10, 23, 29, 30, 6, 9, 7, 34, 35,
8, 16, 33

Faktor III: Angst vor schulischer Lei-
stungsprüfung

TASC: Item Nr. 67, 60, 70, 61, 65, 66, 47, 69,
45, 72, 42

CMAS: Item Nr. 32

Faktor IV: Gefühl der Unterlegenheit (?)

CMAS: Item Nr. 38, 7, 19
TASC: Item Nr. 51, 48 "1).

Die Faktorenstruktur der TASC ist also prägnanter als die der CMAS. Die Prüfungsangstskala ist "in der vorliegenden Form wesentlich eindeutiger und vor allem für die Erfassung schulisch relevanter Aspekte der Angst (Leistungs- und Prüfungsangst) geeigneter als die CMAS"²⁾.

4.5.4. Fremdspracheneignungstest für die Unterstufe: FTU 4 - 6

Das Konstruktionskalkül von LPS, FTU 4 - 6 und den beiden informellen Tests I und II basiert auf den Grundannahmen der klassischen Testtheorie, die kaum Restriktionen enthält. Dadurch ist auch der empirische Gehalt³⁾ der klassischen Testtheorie größer als z.B. der eines logistischen Modells.⁴⁾ "The classical test theory model ... is an example of a model based on relatively weak assumptions and hence one that has wide applicability. Such models may be called weak models (weak true-score models). Other models ... adopt assumptions that are satisfied only under very limited conditions. Such models may be called strong models (strong true-score models)."⁵⁾

1) NICKEL, H. u.a., a.a.O., S. 3

2) NICKEL, H. u.a., a.a.O., S. 7

3) vgl. OPP, K.-D.: Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theoriebildung, Hamburg 1970

4) FRICKE, R.: Über Meßmodelle in der Schulleistungsdiagnostik, Düsseldorf 1972

5) LORD, F.M./NOVICK, M.R.: Statistical Theories of Mental Test Scores, Massachusetts/London/Ontario 1968, S. 25

4.5.4.1. Exkurs (1): Grundannahmen der klassischen Testtheorie

Das Verknüpfungssaxiom $X = T + E$ (1) ¹⁾ behauptet, daß sich der beobachtete Testwert einer Person i aus dem wahren Testwert T und dem Fehleranteil E zusammensetzt. Weitere Zusatzannahmen sind:

- $\sigma E = 0$ (2) Der Erwartungswert (oder: das arithmetische Mittel) aller Fehler E ist null, wenn man der Vp den Test n -mal gibt (Lerneffekte seien hierbei ausgeklammert);
- $\rho_{TE} = 0$ (3) Die Korrelation zwischen 'wahrem Wert' und 'Fehlerwert' ist null (stochastische Unabhängigkeit);
- $\rho(E_1, E_2) = 0$ (4) Die Meßfehler weisen eine Null-Korrelation auf;
- $\rho(E_1, T_2) = 0$ (5) Es besteht keine Korrelation zwischen E_1 und T_2 .

Den Hauptmangel der klassischen Testtheorie erblickt man darin, daß "das Verknüpfungssaxiom ... nicht empirisch prüfbar"²⁾ ist. FISCHER kommt zu dem Schluß: "Die semantische Interpretation des syntaktischen Gerüsts erfordert also die Annahme, daß sich psychologische Messungen (Testungen) ohne Lern-, Transfer- und Ermüdungswirkungen wiederholen lassen."³⁾ SIXTL weist "die faktische Inkonsistenz der Testtheorie"⁴⁾ nach bei der Verwendung des Phi-Koeffizienten, der "nur dann eine erwartungstreue Schätzung des Korrelationskoeffizienten (ist), wenn die Items gleich schwierig sind"⁵⁾. Die 'beobachteten' Itemschwierigkeiten variieren jedoch. Eine weitere Restriktion besteht in der Populationsabhängigkeit, die deutlich an der Formel des Produkt-Moment-Korrelationskoeffizienten zu erkennen ist, in den ja die populationsgebundenen Streuungen der Variablen eingehen. Auch meßtheoretische Argumente werden gegen die klassische Testtheorie vorgebracht.

-
- 1) LORD, F.M. u.a., a.a.O., S. 56
- 2) FRICKE, R.: Zur Theorie des Rasch - Modells und einige Erfahrungen bei der Anwendung dieses Modells auf Schulleistungstests, in: Z e F, 1972, S. 3
- 3) FISCHER, G.H. (Hrsg.): Psychologische Testtheorie, Bern/Stuttgart 1968, S. 24
- 4) SIXTL, F.: Kritik an der theoretischen Konzeption der Leistungstests: die faktische Inkonsistenz der Testtheorie, in: Ingenkamp, K./Marsolek, T.: Möglichkeiten und Grenzen der Testanwendung in der Schule, Weinheim 1968, S. 467 ff. vgl. SIXTL, F.: Beiträge zur Explikation und Kritik der klassischen Testtheorie, in: Fischer, G.H. (Hrsg.), a.a.O., S.
- 5) SIXTL, F. (1968), a.a.O., S. 468

"Der wesentliche Unterschied besteht darin, daß die moderne Theorie schon vor der Messung einsetzt und diese erst begründet, während die klassische Theorie von dem Vorliegen von Messungen ausgeht."¹⁾ Ein vernichtendes Urteil fällt FISCHER: "Die Erkenntnis eines allgemeinen Gesetzes muß der Messung vorausgehen. Wird diese Forderung nicht erfüllt, dann handelt es sich nicht um Messung, sondern nur um Schein-quantifizierung, um Benennung von Ereignissen mit Zahlen anstelle beliebiger anderer Namen."²⁾ GUTJAHR versucht aufgrund eigener Untersuchungsergebnisse zwischen der 'klassischen' und 'modernen' Testtheorie zu vermitteln: "Es wird notwendig sein, an einer Vielfalt empirischer Daten den Vergleich zwischen beiden Modellen zu wiederholen; dabei könnte sich sehr gut zeigen, daß jedes der Modelle das für bestimmte Sachverhalte adäquatere ist."³⁾

Die Kriterien für die Anwendung einer Methode liegen in dem durch die jeweilige Theorien- und Hypothesenbildung konstituierten Sachverhalt. "Ausgangspunkt aller Probleme der quantitativen Untersuchung in der Pädagogik ist daher nicht die technische Frage, welche mathematischen Methoden man verwenden sollte, sondern das viel tiefer liegende Problem, wie man die objektiven quantitativen Charakteristiken der Erscheinungen der Bildung und Erziehung richtig bestimmt."⁴⁾

4.5.4.2. Charakterisierung des FTU 4 - 6

Der FTU 4 - 6 ist die deutsche Adaption der "Elementarform des bewährten Modern Language Aptitude Test von J.B. CARROL und S.M. SAPON, der auf langjährige Grundlagenforschung in den USA beruht."⁵⁾ Mit den vier Faktoren 'phonetic coding', 'grammatical sensitivity', 'rote memorization ability' und 'inductive language learning ability' sind nach CARROL und SAPON wesentliche Kompo-

1) FISCHER, G.H., a.a.O., S. 16

2) FISCHER, G.H., a.a.O., S. 19

3) GUTJAHR, W.: Die Messung psychischer Eigenschaften, Berlin 1961, S. 251

4) ITELSON, L.: Mathematische und kybernetische Methoden in der Pädagogik, Berlin 1967, S. 63

5) FTU 4 - 6, Testmanual, S. 16

nenten der Fremdspracheneignung erfaßt, die auch gute Prognosen über die spätere Fremdspracheneignung erlauben. Dies mag für praktische Zwecke ausreichen, doch kann damit das theoretische Interesse nicht befriedigt werden. "The lack of a construct (von mir hervorgehoben - K.H.) ... makes it very difficult to interpret the factor analysis in terms of anything except combinations of tests. Even so the fact is that the MLAT works."¹⁾ Den vier faktorenanalytisch errechneten Dimensionen kommt - außer einem formalen, der die Reproduktion der manifesten Interkorrelationen gestattet, kein Erklärungswert zu. Teststatistiken und Validität können sich sogar umgekehrt zueinander verhalten, denn "it is possible to imagine a test with very low item discrimination because many heterogeneous abilities are being tested, poor reliability because the range is small; and yet such a test could have high validity, and this is what counts. It is the test constructor's assumptions in language learning that are really being analysed. A good test is a device for framing these assumptions and as much a learning device as a language laboratory or a programmed text."²⁾

Die deutschen Bearbeiter resümieren: "Die hier mitgeteilten Entwicklungsarbeiten und statistischen Kriterien rechtfertigen es, den FTU 4 - 6 als objektives, gültiges und zuverlässiges Verfahren zu bezeichnen."³⁾

4.5.5. Das Leistungs - Prüf - System (LPS) von Horn

Die Variable Intelligenz wurde erhoben, weil eine Wechselwirkung zwischen Angst und Intelligenz zu bestehen scheint. Ihr Anteil beim Erwerb einer Fremdsprache ist weitgehend ungeklärt.⁴⁾

-
- 1) PIMSLEUR, P.: Language Aptitude Testing, in: Davies, A. (ed.): Language Testing Symposium. A Psycholinguistic Approach, London 1968, S. 105
 - 2) DAVIES, A.: Language Testing ..., London 1968, S. 10
 - 3) FTU 4 - 6, Testmanual, S. 3
 - 4) FLECHSIG, K.-H.: Forschung im Bereich des Fremdsprachenunterrichts, in: Ingenkamp, K./Parey, E. (Hrsg.): Handbuch der Unterrichtsforschung, Teil III, Weinheim/Berlin/Basel 1971, Spalte 3239

Das Leistungs-Prüf-System enthält 14 Aufgabenreihen (Subtests). Die Auswahl der items für die Subtests orientierte sich an den THURSTONESchen Intelligenzfaktoren 'verbal factor', 'reasoning', 'wordfluency', 'space', 'closure', 'perceptual speed', 'accuracy' und 'number'. Die Zuordnung der Subtests zu den Faktoren ist folgende:

- "1 + 2 Allgemeinbildung, Rechtschreibkenntnisse (V e r b a l
F a c t o r
- 3 Denkfähigkeit, Erkennen von Gesetzmäßigkeiten, relativ
unabhängig von der Schulbildung (R e a s o n i n g)
- 4 Logisches Denken und Erkennen von Regeln, wobei Schuli-
sches nicht ganz unwichtig ist (R e a s o n i n g)
- 5 + 6 Worteinfall oder Wortflüssigkeit und Initiative (W o r d-
f l u e n c y)
- 7 Gedankliches Bewegen von Symbolen (S p a c e 1)
- 8 Räumliches Vorstellen und Formen mit Symbolvergleich
(S p a c e)
- 9 Raumvorstellung (S p a c e)
- 10 Erkennen des Wesentlichen trotz ablenkender Einzel-
heiten (C l o s u r e)
- 11 Umfang und schnelle Verfügbarkeit visueller Gedächtnis-
vorstellungen (C l o s u r e)
- 12 Erkennen von Unvollständigem (C l o s u r e 1)
- 13+14 Bemerken von Fehlern, Heraussuchen und Zählen von Zahlen,
Wahrnehmungstempo (P e r c e p t u a l S p e e d,
A c c u r a c y)"1)

Die Behauptung des Testautors, das Leistungs-Prüf-System erfasse "die üblichen Schulleistungen nur begrenzt"²⁾, stimmt mit den Untersuchungsergebnissen TENTs überein. "In der Reihenfolge ihres Gewichts läß sich das LPS inhaltlich in folgende sechs D i m e n s i o n e n aufgliedern: (1) Sprachliche Leistungsfähigkeit, (2) Schlußfolgerndes Denken, (3) Räumliches Vorstellen, (4) Leistungsmotiviertheit, (5) Wiedererkennen von Symbolen und (6) Schulisches Lernen."³⁾

Eine neuere Untersuchung 14-jähriger Hauptschüler, Realschüler und Gymnasiasten weist darauf hin, daß die von HORN publizierten Normen nicht mehr ganz zutreffen. "Unsere Ergebnisse lassen es -

-
- 1) + 2) HORN, W.: Leistungsprüfsystem L-P-S, Handanweisung für die Durchführung, Auswertung und Interpretation, Göttingen 1962, S. 24
 - 3) TENT, L.: Die Auslese von Schülern für weiterführende Schulen, Göttingen 1969, S. 106

zumindest für diese Altersstufe - nicht gerechtfertigt erscheinen, bei allen Subtests einen Mittelwert auf der C - Skala von 5 und eine Standardabweichung von 2 vorauszusetzen."¹⁾

4.5.6. Informelle Tests I und II

Die Kritik der herkömmlichen Prüfungs- und Notenpraxis konzentriert sich in vielen Publikationen auf den Nachweis mangelnder Objektivität und Reliabilität, denn "für jedes Schulsystem ist es von ausschlaggebender Bedeutung, ob es über schulische Prüfungs- und Bewertungsverfahren verfügt, die angemessen, objektiv, gültig und zuverlässig erfassen können, in welchem Maße die Lernziele von Individuen und Gruppen erreicht werden, wie die individuelle Leistungskapazität beschaffen ist und wie sie am besten gefördert werden kann. Bei jeder Schulreform muß geprüft werden, ob die vorhandenen Verfahren zur Erreichung der neuen Ziele geeignet sind."²⁾ Objektivität als Eigenschaft eines Prüfungsinstruments bedeutet intersubjektive Übereinstimmung der Beurteiler.³⁾ Eine andere Verwendungsweise von 'objektiv' bedeutet, "daß die Leistungsanforderungen, die für alle gleich und von daher vergleichbar sind, d.h. eine Lokalisierung der gemessenen Leistung in einer größeren Population erlauben. Objektive Leistungsbeurteilung muß, wenn sie gerecht und lernwirksam sein will, auch die Voraussetzungen und Möglichkeiten des einzelnen Schülers in Rechnung ziehen. Die standardisierten Leistungstests können also nur eine Hilfsfunktion haben."⁴⁾

-
- 1) GEHLERT, S./FREYN, R.: LPS-Profile 14-jähriger Schüler, in: Dietrich, G. (Hrsg.): Unterrichtspsychologie der Sekundarstufe. Beiträge zu Forschung, Theorie und Praxis, Donauwörth 1972, S. 107
 - 2) INGENKAMP, K.: Möglichkeiten und Grenzen des Lehrerurteils und der Schultests, in: Roth, H. (Hrsg.): Begabung und Lernen. Deutscher Bildungsrat, Gutachten und Studien der Bildungskommission, Bd. 4, S. 426
 - 3) Objektivität wird hier reduziert auf Intersubjektivität; LIENERT, G.A., a.a.O., S. 13
 - 4) SCHREINER, G.: Sinn und Unsinn der schulischen Leistungsbeurteilung, in: Die Deutsche Schule, 1970, S. 227

Eine Theorie des Prüfens¹⁾, die sowohl individuellen als auch gesellschaftlichen Ansprüchen genügt, existiert nicht. "Es wird naiv angenommen, daß unser Prüfungswesen objektiv, gerecht, einheitlich, allseitig überzeugend und sowohl dem Individuum als auch der Gesellschaft zureichend angemessen ist."²⁾ Soziologisch gesprochen heißt dies: 'Leistungsprinzip' und 'Sozialprinzip' sind nicht zur Deckung gebracht worden.³⁾

4.5.6.1. Exkurs (2): Die Syntax informeller Tests und ihr Verhältnis zur schulischen Sozialisation und Identitätsfindung des Schülers

Sowohl informelle Tests als auch Klassenarbeiten traditioneller Art stellen normbezogene Leistungsmessungen dar. Den Gegensatz dazu bilden kriteriumbezogene Messungen. Normbezogene Messungen "sagen nur, daß ein Schüler mehr oder weniger gut ist als ein anderer; sie sagen aber nicht, wie gut einer von ihnen in bezug auf den betreffenden Stoff ist", während kriteriumbezogene Messungen "die Schülerleistung in Form von erreichten Standardwerten bestimmen, geben daher Auskunft über den Grad, bis zu dem ein bestimmter Schüler den Lehrstoff gemeistert hat - eine Bestimmung, die von der Leistung anderer Schüler völlig unabhängig ist (von mir hervorgehoben - K.H.)."⁴⁾

Man kann hypothetisch annehmen, daß normbezogene Prüfetechniken Bezugssysteme⁵⁾ erzeugen, welche die soziale Wahrnehmung⁶⁾ anderer Personen aufgrund des internalisierten Bezugssystems beeinflussen.

-
- 1) Das leistet auch nicht: SALZMANN, Chr.: Studien zu einer Theorie des Prüfens und Erprobens im Raum der Erziehung, Ratingen 1967
 - 2) ROTH, H.: Einleitung und Überblick, in: Roth, H. (Hrsg.): Begabung und Lernen, a.a.O., S. 48
 - 3) vgl. MÜLLER, M.: Nachwort, in: Sinn und Unsinn des Leistungsprinzips. Ein Symposium, München 1974
 - 4) GLASER, R.: Unterrichtstechnologie und die Messung von Lernergebnissen, in: Issing, L.J. (Hrsg.): Der Programmierte Unterricht in den USA heute, Weinheim/Berlin 1967, S. 45
 - 5) Zum Begriff des Bezugssystems: WITTE, W.: Das Problem der Bezugssysteme, in: Metzger, W. (Hrsg.): Hdb. d. Ps., Bd. I/1, Allgemeine Psychologie, Göttingen 1966, S. 1003 ff.
 - 6) SECORD, P.F./BACKMANN, C.W.: Social Psychology 1964

Wie aus der Analyse der syntaktischen Elemente informeller Tests hervorgeht, ist das Versagen einzelner Schüler konstitutives Merkmal der Konstruktionsprinzipien. Der programmierte Mißerfolg produziert Verhaltensmechanismen, die sich z.B. in der "Abwertung von Aufgaben" oder "der eigenen Person"¹⁾ manifestieren. Normbezogene Messungen, die mit bestimmten Zensierungsmodellen²⁾ gekoppelt sind, lösen unerwünschte pädagogische Nebenwirkungen aus wie z.B. konkurrenthes Verhalten anstelle von Kooperation. Dagegen könnte kriteriumsorientierte Leistungsmessung - diese Hypothese müßte jedoch ebenfalls empirisch überprüft werden - "dazu beitragen, das mit universellen Lernzielen allemal verknüpfte Problem des Leistungsversagens in der Schule wenigstens teilweise abbauen zu helfen. Normativ gesetzte Ziele und institutionalisierte wie personalisierte Leistungssanktionen bewirken bei nicht wenigen Schülern einen verhängnisvollen Zirkel von Überforderung - Leistungsversagen - Beeinträchtigung von Selbstkonzept und sozialem Status - Schulangst - vermehrter Überforderung usw."³⁾

Prüftechniken haben somit einen Anteil an der Sozialisation des Schülers, da sie zur Interiorisierung der Norm 'Leistung' beitragen. Andererseits verzögern sie die Entwicklung von Ich-Identität und haben damit eine dissoziative Funktion. "Ich - Identität wird aufgefaßt als Balance zwischen der Anforderung, zu sein wie die anderen, und der Anforderung, zu sein wie kein anderer - in der horizontalen Dimension gleichzeitig aktualisierbarer Rollen und in der vertikalen Dimension der individuellen Biographie."⁴⁾

-
- 1) HELM, J.: Über den Einfluß affektiver Spannungen auf das Denken, in: Zeitschr. f. Ps., 1954
 - 2) WEIS, V.: Zensierungsmodelle und ihre pädagogischen Konsequenzen, in: Die Deutsche Schule, 1971; z.B. die FEGA-Werte von GAUDE, P./TESCHNER, W.-P.: Objektivierete Leistungsmessung in der Schule. Einsatz informeller Tests im leistungsdifferenzierten Unterricht, Frankfurt/Berlin/München 1971², S. 88; oder die STASIX von BIGLMAIER, F.: Leistungsmessung durch informelle Lehrertests, in: b : e, 1969
 - 3) WEINERT, F.E.: Schule und Beruf als institutionelle Sozialisationsbedingungen, in: Graumann, C.F. (Hrsg.): Sozialpsychologie, Hdb. d. Ps., Bd. III/2, Göttingen 1972, S. 856
 - 4) WELLENDORF, F.: Schulische Sozialisation und Identität. Zur Sozialpsychologie der Schule als Institution, Weinheim/Basel 1974², S. 48

Dabei schließen wir uns der These WELLENDORFs an, "daß eine ungebrochene Darstellung der persönlichen Identität im szenischen Arrangement der Schule nicht oder nur unzulänglich möglich ist, da in die Inszenierung schulischer Bedeutungsgehalte, wie sie in Ritualen (und eines dieser Rituale ist die Klassenarbeit - K.H.) der Schule exemplarisch findet, die Triebimpulse und Affekte der Individuen, insofern sie in infantilen Szenen interpretiert sind, als bewußt kommunizierbare Bedeutungen eingehen."¹⁾

Die besondere Organisationsstruktur der Schule erzeugt Dissoziationsprozesse, "da die schulischen Interaktionsprozesse unter der Bedingung einer noch relativ schwach ausgebildeten Ich-Identität der größeren Gruppe der Interaktionspartner (der Schüler) ablaufen."²⁾ Es liegt daher nahe, zumindest bei einem Teil der Schüler, eine neurotisierende Wirkung des Strukturelements Prüfung im Organisationssystem Schule anzunehmen. "Wir müssen heute das übliche Prüfungswesen als neurotisches Symptom der Schule bezeichnen. Das heißt, daß von der Einrichtung "Schule" eine neurotisierende Wirkung ausgehen muß."³⁾ Elternhaus und Schule differieren in den Sozialisationspraktiken der Kontaktnahme und -befriedigung. Gerade in der Situation Klassenarbeit ist die Restrangierung sozialer Kontaktnahme überdeutlich, denn hier tritt der Lehrer als Agent des bürokratischen Systems auf. "Gegenüber der persönlichen und damit notwendig individualisierten Beziehung der Eltern zu ihren Kindern bedeutet das eine Entpersönlichung: Der Lehrer als Amtsträger verhält sich nicht zu Kindern als Personen mit individuellen Motiven und Eigenarten, sondern zu Schülern, d.h. nur zu den Eigenschaften von Kindern, die von der Schülerrolle beansprucht und gefordert sind."⁴⁾

1) WELLENDORF, F., a.a.O., S. 49

2) WELLENDORF, F., a.a.O., Anmerkung 97, S.49

3) SINGER, K.: Lernhemmung, Psychoanalyse und Schulpädagogik, München 1970, S. 151

4) FÜRSTENAU, P.: Zur Psychoanalyse der Schule als Institution, in: Pädagogisches Zentrum. Veröffentlichungen. Hrsg.: C.-L. Furck. Reihe B: Diskussionsbeiträge. Bd. 10. Zur Theorie der Schule. Mit Beiträgen von P. Fürstenau u.a., Weinheim/Berlin/Basel 1969, S.11. Zum bürokratischen System Schule vgl. im gleichen Band den Beitrag: FÜRSTENAU, P.: Neuere Entwicklungen der Bürokratieforschung und das Schulwesen. Ein organisationssoziologischer Beitrag, S.47ff., vgl. MAYNTZ, R./ZIEGLER, R.: Zur Soziologie der Organisation, in: König, R. (Hrsg.): Hdb. d. empir. Sozialforschung, Bd. II, Stuttgart 1969, S. 444 - 513

Den weiteren Ausführungen sei unsere noch zu belegende These vorangestellt: Die *S y n t a x* informeller Tests (sichtbar in den Konstruktions- und Analyseschritten) erzeugt dissoziative Effekte in der Schülerpsyche.

"Unter einem informellen Test (Lehrertest, classroom test, teacher-made test ...) verstehen wir eine nach bestimmten Kriterien auf objektive Auswertbarkeit angelegte Probearbeit für ausgewählte Gebiete eines Unterrichtsfachs mit definierter Zielsetzung. Der didaktische Ort (Einsatzort), alle antizipierbaren Variablen der Durchführung, der Lösungsschlüssel und das Auswertungsschema (die Zensierung) sind fixiert."¹⁾ Das *K o n s t r u k t i o n s p r i n z i p* informeller Tests ist die *M a x i m i e r u n g* der *V a r i a n z*. "Normorientierte Tests sind so konstruiert, daß sie die Varianz zwischen den getesteten Personen maximieren, um so Unterschiede zwischen den Personen zu entdecken. Für kriteriumsbezogene Messungen ist Variabilität unerheblich; eine Messung kann auch dann gut sein, wenn alle Personen den Maximalpunktwert bekommen."²⁾

Bei der *K o n s t r u k t i o n* informeller Tests werden festgelegt:

- | | |
|---|--------------------------|
| (1) LERNZIEL (operationalisiert; in der Regel dem Schüler nicht bekannt | (4) AUSWERTUNGSSCHLÜSSEL |
| (2) INSTRUKTION (standardisiert) | (5) ZENSURENSCHLÜSSEL |
| (3) AUFGABENFORM (objektiviert | |

Die *T e s t a n a l y s e* ermittelt:

- | | |
|---|--|
| (1) SCHWIERIGKEITSKOEFFIZIENTEN der items | (3) RELIABILITÄTSKOEFFIZIENTEN/STANDARDMESS-FEHLER |
| (2) TRENNSCHÄRFEKOEFFIZIENTEN der items | |
| | (4) NORMEN |

Die umfangreiche Literatur zum Testaufbau und zur Testanalyse³⁾ beschränkt sich - auch wenn die Autoren Pädagogen sind - auf die

1) GAUDE, P./TESCHNER, W.-P., a.a.O., S. 48

2) Weis, V., a.a.O., S. 551

3) BELSER, H.: Testentwicklung. Verfahren und Probleme der Entwicklung von Gruppen-Intelligenztests. Dargestellt am Frankfurter Analogietest, Weinheim 1967
BIGLMAIER, F.: Leistungsmessung durch informelle Lehrertests, in: B : e, 1969

Darstellung der Techniken, so daß wir für die Stützung unserer Hypothese dort keine direkten Lösungshilfen erhielten.

Schon die erste Phase der Testkonstruktion, die Bestimmung der LERNZIELE, wird vom Lehrer ohne Mitarbeit der Schüler geleistet, so daß die anonyme Leistungsforderung angstausslösend wirkt. Hier ist vom pädagogischen Standpunkt aus Transparenz der Lernziele zu fordern.

Fortsetzung der Literaturangaben von der vorhergehenden Seite:

CHAUNCEY, H./DOBBIN, J.E.: Der Test im modernen Bildungswesen, Stuttgart 1968

DIEDERICH, P.B.: Short-cut Statistics for Teacher-made Tests. Educational Testing Service. Evaluation and Advisory Service Series No. 5, Princeton 1960

DRENTH, P.: Der psychologische Test, München 1969

EKMAN, G.: Konstruktion und Standardisierung von Tests, Göttingen 1959

ETHERTON, A.R.B.: Objective English Tests for Secondary Schools, London/Harlow 1969⁴

GRONLUND, N.E.: Measurement and Evaluation in Teaching, New York 1965

HEUSER, J./MESSELKEN, H.: Leistungskontrolle im Fremdsprachenunterricht, in: Probleme, Prioritäten, Perspektiven des fremdsprachlichen Unterrichts. Informationen des Hessischen Instituts für Lehrerfortbildung, Frankfurt am Main 1972

HORST, P.: Messung und Vorhersage. Eine Einführung in die psychologische Testtheorie, Weinheim/Berlin/Basel 1971

INGENKAMP, K.: Die deutschen Schulleistungstests, Weinheim 1962

INGRAM, E.: Appendix. Item Analysis, in: Davies, A. (ed.): Language Testing Symposium. A Psycholinguistic Approach, London 68

KAMRATOWSKI, J./PENNE, K.J./SCHNEIDER, J.: Informelle Fremdsprachentests für die Schulpraxis, Berlin 1970

LADO, R.: Language Testing. The Construction and Use of Foreign Language Tests, London 1967

LIENERT, G.A.: Testaufbau und Testanalyse, Weinheim/Berlin 1967²

LINDQUIST, E.F.: Educational Measurement, Washington 1963⁵

MICHEL, L.: Allgemeine Grundlagen psychometrischer Tests, in: Heiss, R./Groffmann, K.J./Michel, L. (Hrsg.): Psychologische Diagnostik, Hdb. d. Ps., Bd. VI, Göttingen 1964

PIMSLEUR, P.: Testing Foreign Language Learning, in: Valdman, A. (ed.): Trends in Language Teaching, 1966

Die gesellschaftlich vermittelte Präferenz für individuelles privates Arbeiten, die Betonung des Konkurrenzcharakters in Prüfungssituationen findet ihren Ausdruck in der standardisierten INSTRUKTION. "Dabei tritt ein Grundprinzip institutioneller Kontrolle deutlich hervor: das der Reduktion von Interaktion auf Einzelkontakte. Die Schüler dürfen im Hinblick auf das von ihnen erwartete Leistungsverhalten nicht mehr miteinander kommunizieren; andernfalls 'mogeln' sie."¹⁾

Von den AUFGABENFORMEN²⁾ scheint die multiple-choice-Form am meisten bevorzugt zu werden. Konstituierend für diese Aufgabenform sind die Distraktoren, die auf seiten des Schülers sprachliche Kompetenz voraussetzen. Da nun einerseits sprachliche Performanz qualitative und quantitative Unterschiede aufweist, die das Ergebnis einer bestimmten Schichtzugehörigkeit sind, andererseits jedoch die Sprachnorm einer Schicht die sprachliche Form der Distraktoren kanalisiert, werden Schüler der sozialen Schicht,

1) WELLENDORF, F.; a.a.O., S. 121

2) Ob z.B. konvergente Verhaltensweisen zu einem Teil das Ergebnis langjähriger schulischer Erfahrungen mit Aufgabenformen sind, die dem Schüler zwar Probleme stellen, sie ihn aber nicht finden lassen, so daß seine kreativen Fähigkeiten verkümmern, müßte empirisch - trotz großer methodischer Probleme - untersucht werden

Fortsetzung der Literaturangaben von der vorhergehenden Seite:

SAMSTAG, K./BAUS, M.: Pädagogisch-psychologisches Testen. Anleitung zur Testarbeit in der Schule, Bad Heilbrunn 1962

Süllwold, F.: Schultests, in: Heiss, R./Groffmann, K.J./ Michel, L. (Hrsg.): Psychologische Diagnostik, Hdb. d. Ps., Bd. VI, Göttingen 1964

VALETTE, R.M.: Tests im Fremdsprachenunterricht, Berlin 1971

VUKOVICH, A.F.: Die Konstruktion psychologischer Tests, in: Heiss, R./Groffmann, K.J./Michel, L. (Hrsg.): Psychologische Diagnostik, Hdb. d. Ps., Bd. VI, Göttingen 1964

WENDELER, J.: Standardarbeiten - Verfahren zur Objektivierung der Notengebung, Weinheim/Berlin/Basel 1969

die nicht den vom Lehrer erwünschten Code¹⁾ spricht, die Benachteiligten sein.

BEWERTUNGS- und ZENSURENSCHLÜSSEL, die wie z.B. die *S t a s i x* sich an der GAUSSschen Normalverteilungskurve orientieren, begünstigen im schulischen Sozialisationsprozeß die Internalisierung des Leistungsprinzips. "Mit Hilfe des Modells der Normalverteilung werden die symbolischen Gratifikationen, die im Interaktionssystem der Schule für Leistung vergeben werden, behandelt, als wären die Ressourcen begrenzt, als käme es darauf an, jedem Schüler einem seinen Leistungen entsprechenden Anteil an einer begrenzten Menge Gratifikationen zukommen zu lassen - wie im Bereich der Arbeitsorganisationen der einzelne gemäß der Ideologie des Leistungsprinzips einen seiner Leistung entsprechenden Anteil am Sozialprodukt erhalten soll. Dabei unterscheidet sich das Verteilungsmodell der Schule von dem der Organisationen zwar der Form nach, wichtig ist jedoch die symbolische Darstellung einer zentralen Prämisse für die Verteilung des Sozialproduktes in schulischen Leistungsritualen: daß die Ressourcen für Gratifikationen prinzipiell begrenzt seien. Damit wird in den Ritualen zugleich der selektive Charakter schulischer Interaktionsprozesse, durch den Schule und Gesellschaft miteinander verknüpft sind, mit dargestellt: die schulinternen Verteilungsprozesse für symbolische Gratifikationen legen den Anteil, den der einzelne Schüler später außerhalb der Schule am Sozialprodukt erhält, in einer wenn auch groben Klassifi-

-
- 1) BERNSTEIN, B.: Soziokulturelle Determinanten des Lernens, in: Heintz, P. (Hrsg.): *Soziologie der Schule*, Köln/Opladen 1959
OVERMANN, U.: *Sprache und soziale Herkunft. Ein Beitrag zur Analyse schichtenspezifischer Sozialisationsprozesse und ihrer Bedeutung für den Schulerfolg*, Frankfurt am Main 1972
OVERMANN, U.: *Schichtenspezifische Formen des Sprachverhaltens und ihr Einfluß auf die kognitiven Prozesse*, in: Roth, H. (Hrsg.): *Begabung und Lernen*, a.a.O., S. 297 - 356
ROEDER, P.-M.: *Sprache. Sozialstatus und Bildungschancen*, in: Roeder, P.-M./Pasdzierny, A./Wolf, W.: *Sozialstatus und Schulerfolg*, Heidelberg 1965, S. 5 - 32
ROEDER, P.-M.: *Thesen zur Auslese durch den neusprachlichen Unterricht*, in: Roeder, P.-M. (Hrsg.): *Pädagogische Analysen und Reflexionen, Festschrift für Elisabeth Blochmann zum 75. Geburtstag*, Weinheim/Berlin 1967

zierung fest. Denn 'Schulleistung wird- zumindest tendenziell - als ein legitimes Kriterium für die Zuteilung beruflicher, ökonomischer und sozialer Chancen allgemein akzeptiert. Erst auf der Basis der zweiten Funktion, der Knappheit symbolischer Gratifikationen, kann die Gauss'sche Normalverteilung von Zensuren den Individuen in der Schule den Eindruck vermitteln, sie würden 'gerecht' behandelt."¹⁾ Der Konkurrenzkampf um die symbolischen Gratifikationen in Form von Zensuren verhindert die Entwicklung solidarischen Verhaltens und Handelns.²⁾ Außerdem werden hier die Voraussetzungen für den Leistungskonflikt im späteren Berufsleben geschaffen, denn "Berufsrollen werden primär nach Nichtleistungskriterien vergeben. Dennoch stellt Leistung einen zentralen Wert in unserer Gesellschaft dar; sie wird von der Leistungsideologie beherrscht. Die Diskrepanz zwischen sozialer Realität und ideologischem Glauben ergibt den sogenannten Leistungskonflikt: Wir glauben an Leistung als Wert, leben aber in einer Gesellschaft, die sich kaum nach diesem Wert richtet."³⁾ Die Identifizierung mit bestimmten Leistungsnormen ergibt ein weiteres Problem. "Man kann sich durchaus an L e i s t u n g s n o r m e n orientieren, die in keinerlei erkennbarer Beziehung zum tatsächlichen L e i s t u n g v e r h a l t e n stehen. Werthaltungen und Zielsetzungen wurden nämlich verinnerlicht, nicht jedoch die Verhaltensmechanismen, die zum Erreichen derselben erforderlich sind."⁴⁾ Nach dieser Analyse kann man sich nur über die Naivität eines Autors wundern, der zu dem Ergebnis kommt: "So schafft die an der Normalverteilung orientierte Zeugnisgebung ein gesundes pädagogisches Spannungsfeld, das die Kräfte aller Schüler fordert und damit jede Begabungshöhe zur Entfaltung ihrer individuellen Möglichkeiten anspornt."⁵⁾

Es sei schon jetzt vorweggenommen, daß die technischen Schritte, die bei der T e s t a n a l y s e vorgenommen werden, die gleiche Problematik enthalten wie die des T e s t a u f b a u s.

1) WELLENDORF, F., a.a.O., S. 125

2) MOSER, H.: Leistungsprinzip: Ökonomische Kategorie und soziale Norm, in: Die Deutsche Schule, 1972, S. 207

3) SEIBEL, H.D.: Gesellschaft im Leistungskonflikt, Düsseldorf 1973, S. 9

4) GUSKIN, A.E./GUSKIN, S.L.: Sozialpsychologie in Schule und Unterricht, Ulm 1973, S. 114

5) GÖLLER, A.: Zensuren und Zeugnisse, Stuttgart 1966, S. 32

Zwischen TRENNSCHÄRFE und SCHWIERIGKEITSGRAD eines items besteht ein paraboloider Zusammenhang. Die Trennschärfe ist bei einem Schwierigkeitsgrad von $p_i = 0.5$ am größten, weil dann die Varianz ihr Maximum erreicht hat.¹⁾ Das pädagogisch Bedenkliche dieses Ansatzes wird besonders deutlich in der Formel zur Berechnung der Varianz einer Aufgabe i:

$$s_i^2 = p_i (1 - p_i) = p_i \cdot q_i \quad (2)$$

"Die Varianz jeder beliebigen einzelnen Testaufgabe wird also durch das Produkt aus dem Verhältnis getesteter Individuen, die bei der Aufgabe versagen, und aus dem Verhältnis jener, die sie lösen, erhalten."³⁾ Der Kalkül produziert aufgrund seines Mechanismus 'kluge' und 'dumme' Schüler.

In die Formel zur Berechnung des RELIABILITÄTSKOEFFIZIENTEN⁴⁾ geht der gleiche 'Sachverhalt' ein, da auch für sie der Varianzbegriff konstituierend ist:

$$r = 1 - \frac{M(n - M)}{n \cdot s^2}$$

Die Streuungsberechnung verlangt - in den Kurzformeln⁵⁾ kommt dies prägnant zum Ausdruck - 'obere Schüler' (N_o) und 'untere Schüler' (N_u):

$$s = \frac{\frac{1}{6} N_o \cdot RW - \frac{1}{6} N_u \cdot RW}{\frac{1}{2} N}$$

Die crux der NORMIERUNG besteht darin, daß der Beurteilungsmaßstab nicht unabhängig von den zu Beurteilenden gewonnen wird. "Das Ziel jeder Normierung ist es, an einer möglichst repräsentativen Stichprobe von Personen, für die ein bestimmtes Meßverfahren

1) LIENERT, G.A., a.a.O., S. 40

2) Nebenbemerkung: In einem Curriculum zur Lehrerausbildung müßte das Qualifikationsmerkmal 'Verbalisierenkönnen von Formeln' aufgenommen werden; dadurch würden Lehrer nicht von dem objektivistischen Schein statistischer Kalküle geblendet werden.

3) MAGNUSSON, D.: Testtheorie, Wien 1969, S. 27

4) GAUDE, P./TESCHNER, W.-P., a.a.O., S. 120

5) DIEDERICH, P.B.: Short-cut Statistics for Teacher-made Tests. Educational Testing Service. Evaluation and Advisory Series. No. 5, 1964²; deutsch: Diederich, P.B.: Statistische Kurzverfahren zur Analyse informeller Tests, in: Chauncey, H./Dobbin, J.E.: Der Test im modernen Bildungswesen, Stuttgart 1968

konstruiert wurde, Werte zu gewinnen, die als Vergleichsmaßstab für die individuelle Testleistung benutzt werden können."¹⁾

Die Orientierung an der Normalverteilungskurve hat eine politische Komponente: "Da zwangsläufig aufgrund der impliziten 'Normalverteilungs' - Ideologie ständig 'schlechte' Schüler und 'Dumme' produziert werden, damit Auslese statt Förderung gerechtfertigt und die Durchsetzung von Strategien des zielerreichenden Lernens behindert wird, folgt, daß - unabhängig von den Zielsetzungen und Inhalten des Unterrichts - die ohnehin schon Begünstigten, die Kinder der Mittelschicht, weitere Vorteile erhalten, während die ohnehin schon Benachteiligten, die Arbeiterkinder, absolut und relativ jenen gegenüber weiter in Nachteil geraten. Zumindest gegenüber der Klassenlage der Schüler ist das Instrumentarium der 'objektivierten Leistungsmessung' keineswegs neutral."²⁾

4.5.6.2. Beschreibungsskizze der informellen Tests I und II

Eine Beschreibungsmatrix informeller Tests für das Fach Englisch ist unvollständig, wenn sie nicht sprachwissenschaftliche, sprachpsychologische und fachdidaktische Kriterien für die formale Beschreibung von sprachlichen Leistungen enthält.

Leistungsüberprüfung ist abhängig von den vorgängig getroffenen didaktischen Entscheidungen. Das Theorem der sogenannten 'lerntheoretischen' Didaktik³⁾ vom Implikationszusammenhang von Intention, Thematik, Methodik und Medienwahl schränkt aber die

1) GAUDE, P./TESCHNER, W.-P., a.a.O., S. 125 - 126

2) NEANDER, J.: Objektivierte Lernerfolgsmessung in der Gesamtschule - Fortschritt für wen? In: Die Deutsche Schule, 1973, S. 45; vgl. Weis, V., a.a.O., S. 546 und Messner, R.: Die Messung schulischen Erfolgs. Eine Einführung in die Probleme der Kontrolle des Unterrichtserfolgs durch den Lehrer, in: Messner, R./Rumpf, H. (Hrsg.): Didaktische Impulse. Studientexte zur Analyse von Unterricht, Wien 1971, S. 161

3) BLANKERTZ, H.: Theorien und Modelle der Didaktik, München 1969², S. 92 ff.

HEIMANN, P./OTTO, G./SCHULZ, W.: Unterricht - Analyse und Planung, Hannover 1965

Übertragung sprachwissenschaftlicher Kategorien und Theorien¹⁾ auf Lehr- und Lernprozesse ein. Die Beschreibungsmatrix enthält in sich noch keine expliziten Vorstellungen darüber, wie sie (die Sprache - K.H.) wohl zu lehren wäre."²⁾ Diese Problematisierung sprachwissenschaftlicher Erkenntnisse für Unterrichtsentscheidungen trifft analog für die Konstruktion von Leistungstests zu. Wir stimmen daher dem Ansatz MESSELKENS nicht zu: "Den eigentlichen Ansatzpunkt (für eine formale Beschreibungsmatrix - K.H.) kann man dabei nicht im pädagogischen Feld suchen, sondern nur im Sprachwissenschaftlichen selbst - eine reichlich triviale Aussage, die freilich nicht jeder Vertreter der Sprachdidaktik unterschreiben wird."³⁾

Steht im Mittelpunkt der sprachwissenschaftlichen Analyse der Sprachkörper (corpus) als Objekt⁴⁾, dessen strukturelle Komponenten zu identifizieren sind, so beschäftigt sich der Sprachpsychologe mit dem Medialcharakter von Sprache, dem Sprechakt, und es interessiert ihn "nicht so sehr eine Grammatik, welche aus einer Beschreibung von grammatikalischen Sätzen besteht, sondern eine Grammatik, die ein System von Regeln formuliert, nach welchen grammatische Sätze, d.h. akzeptable Morphem-Sequenzen, produziert, gebaut werden können."⁵⁾ Kom-

-
- 1) BARRUTIA, R.: Die neuere Linguistik und ihr methodisches Risiko, in: Freudenstein, R./Gutschow, H. (Hrsg.): Fremdsprachen. Lehren und Erlernen, München 1972
DENNINGHAUS, F.: Die wechselseitigen Einflüsse zwischen der Linguistik und dem Fremdsprachenunterricht, in: Praxis des neusprachlichen Unterrichts, 1971
BÜNTING, K.-D.: Einführung in die Linguistik, Frankfurt 1971
GLINZ, H.: Linguistische Grundbegriffe und Methodenüberblick, Frankfurt/M. 1971
GOTZ, D./BURGSCHMIDT, E.: Einführung in die Sprachwissenschaft für Anglisten, München 1971
HULLEN, W.: Linguistik und Englischunterricht, Heidelberg 1971
LYONS, J.: Einführung in die moderne Linguistik, München 1971
PETERSEN, H.: Reformbestrebungen im Englischunterricht und ihre Kriterien, in: Linguistik und Didaktik, 1970
PETERSEN, H. (1971): Überlegungen zur Anwendung der Transformationsgrammatik im Englischunterricht an deutschen Schulen, in: Der fremdsprachliche Unterricht, 1971
2) PETERSEN, H. (1971), a.a.O., S. 28
3) MESSELKEN, H.: Empirische Sprachdidaktik, Heidelberg 1971, S. 21
4) MESSELKEN, H., a.a.O., S. 29
5) HÖRMANN, H.: Psychologie der Sprache, Berlin/Heidelberg/New York 1967, S. 49

munikation im "empraktischen und synsemantischen Umfeld"¹⁾ beansprucht verschiedene psychische Funktionen: rezeptive-produktive, imitative-kreative, denen "nach FRIES Hören und Lesen (rezeptiv) sowie Sprechen und Schreiben (produktiv) als elementare Fertigkeiten bei sprachlichen Vollzügen"²⁾ zugeordnet werden können:

	1. rezeptiv	2. produktiv
1. auditiv	1.1. hören	1.2. sprechen
2. visuell	2.1. lesen	2.2. schreiben" ³⁾

Die Kombination psychologischer und didaktischer Begriffe reicht für die Konstruktion von Testitems jedoch nicht aus. Dies leistet die BLOOMsche Taxonomie⁴⁾, obwohl nicht alle Kombinationen für den Testaufbau relevant sind. Denn die Itemkonstruktion wird vom allgemeinen Lernziel 'Kommunikation' gesteuert. Noch mehr eingengt wird sie durch die Intention, mündliche Kommunikation in den Mittelpunkt des fremdsprachlichen Unterrichts zu stellen. "Es scheint in der Fremdsprachendidaktik unbestritten, daß in den ersten Jahren die mündliche Sprachäußerung den Vorrang vor der schriftlichen hat."⁵⁾ "Das Hauptziel muß der Erwerb sicherer Sprechgewohnheiten in einem sehr eng umschriebenen Bereich sein, der formal die grundlegenden Satzbaumuster und die zu ihrer Anwendung notwendigen Kenntnisse und Fertigkeiten umfaßt, material die häufigsten Situationen des täglichen Lebens deckt."⁶⁾

Sprachwissenschaftliche, psychologische und didaktische Überlegungen bei der Konstruktion informeller Tests für den Fremdsprachenunterricht werden im Ansatz von HEUER/MESSELKEN⁷⁾ berücksichtigt. Nach ihren Vorschlägen konstruierten wir unsere Tests.

	5. Schuljahr		6. Schuljahr	
Kriterium	Subtest	Anz. d.i.	Subtest	Anz. d. i.
Lautdiskrimination	I	16		
Hörverständnis	II, III	10, 10		
Lexeme	IV, V, VI	16, 4, 6	I, II	16, 9
	VII	10	III, IV	5, 8
Syntax	VIII, IX, X	10, 10, 8	V, VI, VII	5, 8, 8

1) MESSELKEN, H., a.a.O., S. 24; 2) S. 29; 3) S. 30

4) BLOOM, B.S. (ed.): Taxonomy of Educational Objectives, Handbook I: Cognitive Domain, New York 1967

5) HEUSER, I./MESSELKEN, H.: Leistungskontrolle im Fremdsprachenunterricht, in: Probleme, ... , Frankfurt am Main 1972

6) GUTSCHOW, H.: Englisch an Volksschulen. Probleme und Arbeitsformen, Berlin/Bielefeld 1964, S. 167

7) HEUSER, I./MESSELKEN, H., a.a.O., S. 231 ff.

5. UNTERSUCHUNGSERGEBNISSE

Prüfungen sind konstitutive Elemente des pädagogischen Feldes und erfüllen eine gesellschaftliche Funktion. Dabei schließen wir uns der Auffassung MOLLENHAUERS an, "daß das pädagogische Feld als Sinnzusammenhang, konkret als Rahmen historisch-gesellschaftlicher Bedingungen für Lernvorgänge gedacht werden muß, daß ein pädagogischer Sinnzusammenhang also nicht aus sich selbst, sondern nur im Kontext seines gesellschaftlichen Ortes bestimmt werden kann."¹⁾

Die Propagierung informeller Tests verläuft parallel zu den Technologisierungstendenzen im Unterricht, sie steht aber im auffälligen Kontrast zu den fehlenden empirischen Untersuchungen, die sich mit den Auswirkungen des Kalküls auf den Schüler beschäftigen und die pädagogischen Nebenwirkungen in die Perspektive aufnehmen. Hier zeigen sich die Grenzen einer positivistisch ausgerichteten Pädagogik, die sich vor allem auf Mittel-Zweck-Relationen richtet. Dabei sind wir der Meinung, daß die Nebeneffekte durchaus mit dem empirisch-positivistischen Methodenarsenal erfaßt werden können. "Im Schulsystem werden jedoch Prüfungsformen - wie standardisierte Fragebogen - ohne jegliche Untersuchung von pädagogischen 'Nebenwirkungen' eingeführt. Dagegen werden die technischen Problem der Testkonstruktion, die Ermittlung von Zensuren, elektronische Datenverarbeitung etc. sehr ausführlich behandelt."²⁾

Die einschlägige Literatur bietet nichts außer Mutmaßungen vom grünen Tisch, enthält Behauptungen, die zwar im Bereich des empirisch Beweisbaren liegen, aber empirisch nicht überprüft worden sind. "Die Auswirkungen von Prüfungen (Tests) wurden in Deutschland noch nicht systematisch untersucht. Berichte darüber stellen hauptsächlich Spekulationen oder Verallgemeinerungen von Einzelbeobachtungen dar."³⁾

1) MOLLENHAUER, K.: Theorien zum Erziehungsprozeß. Eine Einführung in erziehungswissenschaftliche Fragestellungen, München 1974², S. 32

2) KVALE, ST.: Prüfung und Herrschaft. Hochschulprüfungen zwischen Ritual und Rationalisierung, Weinheim/Basel 1972, S. 209

3) KRUMM, V.: Artikel 'Leistung - Schulleistung - Schulleistungsmessung', in: Wulf, Chr.: Wörterbuch der Erziehung, München/Basel 1974, S. 386

Der Stellenwert unserer Untersuchungsergebnisse wird deutlich, wenn man folgende Einschränkungen beachtet:

1. Die Effekte gehen zu Lasten der experimentellen Bedingungen; Repräsentativität der Ergebnisse, welche eine Zufallsstichprobenentnahme voraussetzt, ist nicht gegeben.
2. Die Effekte der experimentellen Bedingungen werden vorsichtig als Wirkungen momentaner Einstellungen angenommen (s. Kapitel 4.5. , S. 56).
3. Die Begriffe 'integrativ' und 'dissoziativ' haben durch die Operationalisierung einen so großen Bedeutungsverlust erlitten, daß sie mit dem Bedeutungsgehalt, den die Analyse des gesellschaftlichen Kontextes erbrachte, inkomparabel sind.
4. Die mit den Methoden der inferenten Statistik gewonnenen Ergebnisse haben wahrscheinlich nur eine deskriptive Funktion, da die Modellprämissen nicht erfüllt zu sein scheinen (z. B. Annahme der Äquidistanz der Intervalle beim Semantischen Differential).

5.1. Stützung der These: Die Variation der Merkmalsausprägungen der abhängigen Variablen ist ein Effekt der experimentellen Bedingungen

Wesentliches Unterscheidungsmerkmal der normbezogenen Prüfinstrumente 'Klassenarbeit' und 'informeller Test' (Klassentest) ist ihr Strukturierungsgrad, der sich am deutlichsten in der Aufgabenform widerspiegelt. Der informelle Test hat eine prägnantere Figur als die herkömmliche Klassenarbeit. Die Wahrnehmungspsychologie hat dafür den Terminus Figur-Grund-Problem geprägt. Ein gesichertes Forschungsergebnis ist, "daß eine Differenzierung nach Figur und Grund unser g e s a m t e s Wahrnehmungserleben charakterisiert"¹⁾. "Das Figur - Grund - Prinzip hat neben der optischen auch für die akustische Wahrnehmung, für die Motorik (z.B. Rechts - Links - Händigkeit), die Gefühle und nicht zuletzt (hier sogar mit besonderen Erfolgen) für das Denken Bedeutung."²⁾

Da nun informelle Tests aufgrund der Konstruktionsregeln eine prägnantere Gestalt erhalten als herkömmliche Klassenarbeiten, berechtigt das zur Ableitung der Hypothese:

Die Übereinstimmung (Korrelation) der Statistiken, die zu Lasten der experimentellen Bedingung 'informeller Test' gehen, muß größer sein als die Übereinstimmung der Statistiken, die zu Lasten der experimentellen Bedingung 'Klassenarbeit' gehen.

1) KRECH, D./CRUTCHFIELD, R.S.: Grundlagen der Psychologie, Bd. I, Weinheim/Berlin/Basel 1969², S. 77

2) DORSCH, F.: Psychologisches Wörterbuch, Hamburg/Bern 1963⁷, S.118

Die größere Prägnanz einer Bedingung müßte sich in einer geringeren Streuung niederschlagen. Ein Streuungsvergleich kann aber nur vorgenommen werden, wenn "die Messung des Merkmals in beiden Stichproben nach gleicher Methodik und in gleicher Skalierung erfolgte und zu in ihrer Größenordnung übereinstimmenden Mittelwerten führte."¹⁾

Wie jedoch die Tabelle 16 zeigt, stimmen die Mittelwerte nicht überein. Um dennoch einen Streuungsvergleich durchführen zu können

Tabelle 16: Mittelwerte und Standardabweichungen der Dimensionen Bewertung (B), Stärke (S) und Aktivität (A) der Konzepte IX - XVI (Selbstbild, Wunschbild, schwacher Schüler, guter Schüler, Englischlehrer, Klassenarbeit, Zensur und Klassentest) der Versuchsgruppe 5. Schuljahr unter den experimentellen Bedingungen 'Klassenarbeit' (KA-5) und 'Klassentest' (KT-5)

	IX (SELBSTBILD)			X (WUNSCHBILD)		
	B	S	A	B	S	A
KA $\bar{x}$	5.360	3.930	3.710	5.780	3.976	3.921
s	0.812	1.772	1.248	0.847	1.655	1.680
KT $\bar{x}$	5.190	3.790	3.640	6.227	3.657	3.790
s	0.878	1.197	1.369	0.844	1.752	1.593
	XI (SCHW. SCH.)			XII (GUTER SCH.)		
	B	S	A	B	S	A
KA $\bar{x}$	4.160	4.437	4.490	5.940	4.274	4.064
s	1.399	1.445	1.543	1.204	1.576	1.563
KT $\bar{x}$	3.291	3.596	3.830	5.952	4.230	4.217
s	1.309	1.498	1.583	0.833	1.798	1.842
	XIII (ENGL.-L.)			XIV (KL.-ARBEIT)		
	B	S	A	B	S	A
KA $\bar{x}$	4.608	3.227	3.521	4.930	3.714	4.064
s	1.688	1.417	1.601	1.392	1.457	1.250
KT $\bar{x}$	5.031	3.284	3.287	4.626	3.750	3.980
s	1.416	1.179	1.340	1.260	1.213	1.347
	XV (ZENSUR)			XVI (KL.-TEST)		
	B	S	A	B	S	A
KA $\bar{x}$	4.784	4.090	4.160	4.752	4.017	4.148
s	1.315	1.403	1.158	1.421	1.331	1.414
KT $\bar{x}$	4.791	3.793	3.997	4.975	3.917	4.090
s	1.294	1.239	1.183	1.312	1.345	1.198

1) CLAUSS, G./EBNER, H.: Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen, Berlin 1968, S. 88

nen, "müssen wir den Streuungswert in Prozenten des arithmetischen Mittels ausdrücken. Diesem Zweck dient der Variationskoeffizient (Variationskoeffizient) V nach Pearson ... $V = 100 \frac{s}{\bar{x}}$."¹⁾

Die aufgerundeten Werte wurden anschließend nach dem Produkt - Moment - Korrelationskoeffizienten korreliert.

Tabelle 17: Mittelwerte und Standardabweichungen der Dimensionen Bewertung (B), Stärke (S) und Aktivität (A) der Konzepte IX - XVI (Selbstbild, Wunschbild, schwacher Schüler, guter Schüler, Englischlehrer, Klassenarbeit, Zensur und Klassentest) der Versuchsgruppe 6. Schuljahr unter den experimentellen Bedingungen 'Klassenarbeit' (KA-6) und 'Klassentest' (KT-6)

	IX (SELBSTBILD)			X (WUNSCHBILD)		
	B	S	A	B	S	A
KA $\bar{x}$	5.456	4.410	3.643	5.760	4.397	4.050
s	0.776	1.238	1.375	0.989	1.536	1.460
KT $\bar{x}$	5.306	4.123	3.003	6.043	3.717	3.270
s	0.883	1.212	1.238	0.746	1.400	1.188
	XI (SCHW. SCH.)			XII (GUTER SCH.)		
	B	S	A	B	S	A
KA $\bar{x}$	4.010	3.764	3.917	5.851	4.274	3.710
s	0.893	1.264	1.320	0.806	1.433	1.417
KT $\bar{x}$	4.070	3.451	4.037	5.560	3.937	3.710
s	0.998	1.168	1.406	0.868	1.257	1.200
	XIII (ENGL.-L.)			XIV (KL.-ARBEIT)		
	B	S	A	B	S	A
KA $\bar{x}$	5.417	4.737	3.703	4.439	3.850	4.103
s	1.243	1.143	1.055	1.113	1.014	0.906
KT $\bar{x}$	5.257	4.103	3.384	4.360	3.704	3.844
s	1.179	1.343	1.092	1.014	0.919	1.032
	XV (ZENSUR)			XVI (KL.-TEST)		
	B	S	A	B	S	A
KA $\bar{x}$	4.650	3.697	3.897	4.757	3.941	3.694
s	1.389	1.224	0.967	1.321	0.893	1.065
KT $\bar{x}$	4.677	3.770	3.911	4.771	3.817	3.964
s	1.098	0.907	0.835	1.012	0.838	0.966

1) CLAUSS, G./EBNER, H., a.a.O., S. 88

Tabelle 18: Variabilitätskoeffizienten: $V = 100/\bar{x}$

Dimensi- onen der Konzepte		5. Schuljahr		6. Schuljahr	
		KA	KT	KA	KT
IX	B	15.15	16.92	14.22	16.64
	S	45.09	31.58	28.07	29.40
	A	33.64	37.61	37.74	41.23
X	B	14.65	13.55	17.60	12.34
	S	41.62	47.91	34.93	37.66
	A	42.85	42.03	36.05	36.33
XI	B	33.63	39.78	22.27	24.52
	S	32.57	41.66	33.58	33.85
	A	34.37	41.33	33.70	34.83
XII	B	21.15	14.00	13.78	15.61
	S	36.87	42.51	33.53	31.93
	A	38.46	43.63	33.19	32.35
XIII	B	36.63	28.15	22.95	22.43
	S	43.91	35.93	24.65	32.73
	A	45.47	40.77	28.49	32.27
XIV	B	29.60	27.24	25.07	23.26
	S	39.23	32.35	26.34	24.81
	A	30.76	33.34	22.08	26.85
XV	B	27.49	27.01	29.87	23.48
	S	34.30	32.67	33.11	24.06
	A	27.84	29.60	24.31	21.35
XVI	B	29.90	26.37	27.77	21.21
	S	33.13	34.34	22.66	21.95
	A	34.09	29.29	28.83	24.37

Die Korrelation der Variabilitätskoeffizienten bestätigt die Vermutung, daß die experimentelle Bedingung 'Klassentest' aufgrund ihres Strukturierungsgrades eine höhere Übereinstimmung zwischen beiden Gruppen zur Folge hat als die Bedingung 'Klassenarbeit': $r_{KA-5/KA-6} = 0.610$ und $r_{KT-5/KT-6} = 0.882$.

Tabelle 19: Nach den Variabilitätskoeffizienten berechnete Korrelationen der experimentellen Bedingungen 'Klassenarbeit' und 'Klassentest'

	KA-6	KT-6
KA-5	$r = 0.610$	$r = 0.740$
KT-5	$r = 0.787$	$r = 0.882$

5.2. Schüler des 5. Schuljahres unterscheiden sich signifikant in der Beurteilung der Konzepte IX - XVI von den Schülern des 6. Schuljahrs

Die Hypothese, daß sich einerseits die Versuchsgruppen KA-5 und KA-6, andererseits KT-5 und KT-6 signifikant voneinander unterscheiden, bezog sich (s.S. 43) auf die Dimension Bewertung, da unserer Meinung nach der Forschungsstand keine Voraussagen über die Dimensionen Aktivität und Stärke zuließ. 'Vermutungen' über die letztgenannten beiden Dimensionen bezeichneten wir als 'offene Hypothesen'; Zusammenhänge können dabei erst nach dem Experiment erkannt werden. Die Auswertung wird sich in diesem Kapitel jedoch auf alle drei Dimensionen beziehen.

Für die Signifikanzprüfung benutzten wir den t-Test. Eine notwendige Bedingung für seine Anwendung ist die Homogenitätsprüfung der Varianzen. Wie die Tabelle 20 zeigt, sind 73 von 96 Varianzen homogen, da sie kleiner sind als der kritische Bereich, so daß die Nullhypothese ($H_0: s^2 = s_1^2 = s_2^2$) angenommen werden kann.

Im Anschluß an die Homogenitätsprüfung wurde die Streuung der Gesamtheit geschätzt. Die Differenzen der Mittelwerte, die auf Signifikanz geprüft werden sollten, wurden aus den Tabellen 16 und 17 errechnet. Die Nullhypothese wurde angenommen, wenn $t < t_{0.05;98} = 1.99$, die Alternativhypothese, wenn der absolute Wert t größer als die Prüfgröße war.

Bei einem Blick auf die t-Werte der Tabelle 20, die ohne Berücksichtigung des Ausprägungsgrades 'Ängstlichkeit' gewonnen worden sind, fällt auf, daß alle signifikanten Unterschiede im Bereich der 'personbezogenen' Konzepte IX - XII zu finden sind.

Tabelle 20: Prüfung der Stichprobenvarianzen auf Homogenität:

$F = s_1^2 / s_2^2$, geschätzte Streuungen der Grundgesamtheit und t - Werte der Gruppen KA-5/K A-6, KA-5/KT-6, KT-5/K A-6 und KT-5/KT-6 der Konzepte IX-XVI

Experim. Gruppe	Konzept	F-Wert	Signifikanz	s(gesch.)	t-Wert	Signifikanz
KA-5/ KA-6	IX: B	1.09	nein	0.794	0.60	nein
	S	2.05	ja	1.528	1.57	nein
	A	0.64	nein	1.313	0.26	nein
KA-5/ KT-6	IX: B	1.21	nein	0.854	0.29	nein
	S	2.14	ja	1.518	0.63	nein
	A	1.02	nein	1.243	2.86	ja
KT-5/ KA-6	IX: B	1.28	nein	0.829	1.63	nein
	S	1.07	nein	1.218	2.55	ja
	A	1.01	nein	1.372	-	-
KT-5/ KT-6	IX: B	1.01	nein	0.880	0.66	nein
	S	1.03	nein	1.204	1.38	nein
	A	1.22	nein	1.305	2.44	ja
KA-5/ KA-6	X: B	1.36	nein	0.920	0.11	nein
	S	1.16	nein	1.462	1.44	nein
	A	1.32	nein	1.573	0.41	nein
KA-5/ KT-6	X: B	1.29	nein	0.798	1.63	nein
	S	1.40	nein	1.533	0.85	nein
	A	2.00	ja	1.455	2.23	ja
KT-5/ KA-6	X: B	1.37	nein	0.919	2.56	ja
	S	1.30	nein	1.648	2.25	ja
	A	1.19	nein	1.528	0.85	nein
KT-5/ KT-6	X: B	1.28	nein	0.796	1.16	nein
	S	1.57	nein	1.585	0.19	nein
	A	1.80	nein	1.405	1.85	nein
KA-5/ KA-6	XI: B	2.46	ja	1.173	0.64	nein
	S	1.31	nein	1.357	2.48	ja
	A	1.37	nein	1.435	2.00	ja
KA-5/ KT-6	XI: B	1.97	ja	1.215	0.37	nein
	S	1.53	nein	1.314	3.77	ja
	A	1.20	nein	1.476	1.52	nein
KT-5/ KA-6	B	2.15	ja	1.120	3.25	ja
	S	1.40	nein	1.386	0.58	nein
	A	1.44	nein	1.457	0.35	nein

Tabelle 20: Fortsetzung

Experim. Gruppe	Konzept	F-Wert	Signifikanz	s(gesch.)	t-Wert	Signifikanz
KT-5/ KT-6	XI: B	1.72	nein	1.163	3.35	ja
	S	1.65	nein	1.343	0.54	nein
	A	1.27	nein	1.497	0.69	nein
KA-5/ KA-6	XII: B	2.23	ja	1.024	0.77	nein
	S	1.21	nein	1.506	-	-
	A	1.22	nein	1.491	1.19	nein
KA-5/ KT-6	XII: B	1.93	nein	1.050	0.62	nein
	S	1.57	nein	1.425	1.16	nein
	A	1.70	nein	1.393	1.26	nein
KT-5/ KA-6	XII: B	1.07	nein	0.820	0.61	nein
	S	1.57	nein	1.626	0.12	nein
	A	1.69	nein	1.643	1.55	nein
KT-5/ KT-6	XII: B	1.09	nein	0.850	2.31	ja
	S	2.05	ja	1.551	0.94	nein
	A	2.36	ja	1.554	1.63	nein
KA-5/ KA-6	XIII	1.84	nein	1.482	2.73	ja
	S	1.54	nein	1.287	2.48	ja
	A	2.30	ja	1.555	0.67	nein
KA-5/ KT-6	XIII	2.05	ja	1.456	1.20	nein
	S	1.11	nein	1.581	3.15	ja
	A	2.15	ja	1.370	0.51	nein
KT-5/ KA-6	XIII	1.30	nein	1.332	1.46	nein
	S	1.07	nein	1.161	5.86	ja
	A	1.61	nein	1.206	1.70	nein
KT-5/ KT-6	XIII	1.44	nein	1.302	0.87	nein
	S	1.30	nein	1.263	3.24	ja
	A	1.51	nein	1.222	0.40	nein
KA-5/ KA-6	XIV: B	1.56	nein	1.260	1.01	nein
	S	2.06	ja	1.255	0.54	nein
	A	1.90	nein	1.091	0.18	nein
KA-5/ KT-6	XIV: B	1.89	nein	1.218	1.35	nein
	S	2.51	ja	1.218	0.04	nein
	A	1.47	nein	1.146	0.96	nein
KT-5/ KA-6	XIV: B	1.28	nein	1.189	0.80	nein
	S	1.43	nein	1.118	0.45	nein
	A	2.21	ja	1.148	0.52	nein

Tabelle 20: Fortsetzung

Experim. Gruppe	Kon- zept	F-Wert	Signifi- kanz	s(gesch.)	t-Wert	Signifi- kanz
KT-5/ KT-6	XIV: B S A	1.54 1.74 1.70	nein nein nein	1.143 1.076 1.199	1.16 0.21 0.57	nein nein nein
KA-5/ KA-6	XV: B S A	1.12 1.31 1.43	nein nein nein	1.332 1.316 1.066	0.40 1.49 1.23	nein nein nein
KA-5/ KT-6	XV: B S A	1.43 2.39 1.92	nein ja nein	1.211 1.181 1.009	0.41 1.35 1.24	nein nein nein
KT-5/ KA-6	XV: B S A	1.15 1.02 1.50	nein nein nein	1.342 1.231 1.080	0.52 0.37 0.40	nein nein nein
KT-5/ KT-6	XV: B S A	1.39 1.87 2.01	nein nein ja	1.200 1.085 1.023	0.39 0.11 0.42	nein nein nein
KA-5/ KA-6	XVI: B S A	1.16 2.22 1.76	nein ja nein	1.371 1.133 1.251	0.02 0.34 1.31	nein nein nein
KA-5/ KT-6	XVI: B S A	1.97 2.52 2.14	ja ja ja	1.233 1.112 1.211	0.08 0.90 0.78	nein nein nein
KT-5/ S A	XVI: B S A	1.01 2.27 1.27	nein ja nein	1.316 1.141 1.133	0.17 0.09 1.77	nein nein nein
KT-5/ KT-6	XVI: B S A	1.68 2.58 1.54	nein ja nein	1.171 1.120 1.088	0.87 0.45 0.58	nein nein nein

5.2.1. Einstufung des Konzepts 'SELBSTBILD'

Die arithmetischen Mittel der Einstufungen auf der Dimension B des Konzepts 'Selbstbild' liegen bei allen Gruppen im positiven Bereich der Skala (s. Grafik 2). Alle Gruppen schätzen sich 'fröhlich', 'ehrlich', 'klug' und 'angenehm' ein. Wie jedoch die Tabelle 21 zeigt, liegen die Unterschiede innerhalb zufälliger Variabilität.

Tabelle 21: Konzept IX: t-Werte der Dimensionen Bewertung, Stärke und Aktivität der Gruppen KA-5/KA-6, KA-5/KT-6, KT-5/KA-6 und KT-5/KT-6

	KA-6			KT-6		
	B	S	A	B	S	A
KA-5	0.60	1.57	0.26	0.29	0.63	2.80 s
KT-5	1.63	2.55s	-	0.60	1.38	2.44 s

Der Tendenz nach negativ sind die Einstellungen der Schüler des 5. Jahrgangs auf der Dimension S (KA-5: $\bar{x} = 3.9$ und KT-5: $\bar{x} = 3.8$). Ihre Einstellungen werden durch die Attribute 'hart', 'stark' und 'kraftvoll' beschrieben. Dagegen stufen sich die Gruppen KA-6 ($\bar{x} = 4.4$) und KT-6 ($\bar{x} = 4.1$) auf der positiven Skalseite mit den Merkmalen 'weich', 'nachgiebig' und 'zart' ein. Signifikant ist der Unterschied zwischen der Gruppe KT-5, die sich auf der negativen Seite der Skala einordnet, und der Gruppe KA-6, deren Einschätzungen sich im positiven Skalenbereich befinden.

Auf der Dimension A unterscheiden sich die Gruppen KA-5 ($\bar{x} = 3.7$) und KT-5 ($\bar{x} = 3.6$) überzufällig von den Schülern des 6. Jahrgangs, die ihre Einstufungen nach dem Einfluß der unabhängigen Variablen 'Klassentest' vorgenommen haben. Die Zügigkeit der Aufgabendarbietung bei der Administration des informellen Tests scheint die Präferenz für die Eigenschaften 'eilig', 'lebhaft' und 'kraftvoll' ausgelöst zu haben.

5.2.2. Einstufung des Konzepts 'WUNSCHBILD'

Erwartungsgemäß fallen auf der Bewertungsdimension die Mittelwerte des Konzepts 'Wunschbild' höher aus als die des Realbildes. Man möchte freundlicher und fröhlicher sein, als man in Wirklichkeit ist. Nur ein Unterschied erweist sich dabei als signifikant: Die Schüler der Gruppe KA-6 ($\bar{x} = 5.8$) stufen sich auf der positiven Skalseite weniger hoch ein als die der Gruppe KT-5 ($\bar{x} = 6.2$).

Tabelle 22: Konzept X: t-Werte der Dimensionen Bewertung, Stärke und Aktivität der Gruppen KA-5/KA-6, KA-5/KT-6, KT-5/KA-6 und KT-5/KT-6

	KA-6			KT-6		
	B	S	A	B	S	A
KA-5	0.11	1.44	0.41	1.63	0.85	2.23 s
KT-5	2.56s	2.25s	0.85	1.16	0.19	1.85

Die beiden zuletzt genannten Gruppen diskriminieren auch auf der Dimension S signifikant zwischen den polaren Adjektivpaaren 'weich-hart', 'nachgiebig - stark' und 'zart - kraftvoll'. Die Schüler des 5. Schuljahrs, die der experimentellen Bedingung 'Klassentest' unterworfen worden sind (KT-5: $\bar{x} = 3.7$), möchten lieber 'hart', 'stark' und 'kraftvoll' sein. Dagegen wünscht sich die Gruppe KA-6 ($\bar{x} = 4.4$) die Eigenschaften 'weich', 'nachgiebig' und 'zart'. Zieht man zum inferenten Vergleich die Mittelwerte der Gruppen KT-6 ($\bar{x} = 3.7$) und KA-6 ($\bar{x} = 4.4$) heran, so erhält man einen signifikanten t-Wert ($t = 2.38$). Es kann die Interpretation gewagt werden, daß die Bedingung 'Klassentest' das ideale Selbstkonzept auf die negative Skalenseite projiziert. Die Schüler ordnen ihrem Wunschbild die Eigenschaften 'hart', 'stark' und 'kraftvoll' zu, also Merkmalen, die dem Faktor Aggressivität nahekommen.

Auch auf der Dimension A finden wir bei den Gruppen KA-6 ($\bar{x} = 4.1$) und KT-6 ($\bar{x} = 3.3$) die gleiche Tendenz: Die Bedingung 'Klassenarbeit' verursacht Einstufungen auf der positiven Skalenseite, die Bedingung 'Klassentest' lenkt die Wunschvorstellungen in die Richtung des negativen Pols. Der t-Wert ($t = 2.93$) ist signifikant. Statistisch gesichert ist auch der Unterschied zwischen den Gruppen KA-5 ($\bar{x} = 3.9$) und KT-6 ($\bar{x} = 3.3$): $t = 2.23$.

Im Gegensatz zur Versuchsgruppe 6tes Schuljahr üben die unterschiedlichen experimentellen Bedingungen zwischen den Gruppen KA-5 ($\bar{x} = 3.9$) und KT-5 ($\bar{x} = 3.8$) keinen statistisch gesicherten Effekt aus. Bei den Schülern des 6. Schuljahrs hatten wir aufgrund der Differenzierungshypothese - ältere Schüler nehmen differenzierter wahr - einen Effekt angenommen. Dieser ist auch eingetroffen. Die Bedingung 'Klassentest' determiniert die Einstufungen der Schüler des 6. Schuljahrs in die mit den Adjektiven 'eilig', 'rege' und 'lebhaft' beschriebene Richtung.

5.2.3. Einstufung des Konzepts 'SCHWACHER SCHÜLER'

Die Beurteilungen des Konzepts XI durch die Gruppen KT-5 ($\bar{x} = 3.3$) und KA-5 ($\bar{x} = 4.2$) auf der Dimension B weisen eine überzufällige Differenz der Mittelwerte auf. Nach der Administration des informellen Tests halten die Schüler des 5. Jahrgangs den leistungs-

schwachen Schüler für 'traurig', 'unfreundlich', 'unehrlich', 'schlecht', 'dumm' und 'unangenehm'. Anders verhält sich die Gruppe KA-5. Sie gesteht ihm gerade noch ($\bar{x} = 4.2$) etwas Fröhlichkeit, Freundlichkeit usw. zu. In den gleichen Zusammenhang kann auch der signifikante t-Wert der Gruppen KT-5 ($\bar{x} = 3.3$) und KA-6 ($\bar{x} = 4.0$) eingeordnet werden. Das Konzept 'schwacher Schüler' wird von den älteren Schülern (KT-6: $\bar{x} = 4.1$) nach der gleichen Versuchsbedingung signifikant negativer eingeschätzt als von den jüngeren (KT-5: $\bar{x} = 3.3$). Bis zur Falsifikation durch weitere Untersuchungen halten wir an der Hypothese fest, daß informelle Tests, besonders bei jüngeren Schülern, eher eine negative Besetzung des Stereotyps 'schwacher Schüler' begünstigen.

Tabelle 23: Konzept XI: t-Werte der Dimensionen Bewertung, Stärke und Aktivität der Gruppen KA-5/KA-6, KA-5/KT-6, KT-5/KA-6 und KT-5/KT-6

	KA-6			KT-6		
	B	S	A	B	S	A
KA-5	0.64	2.48s	2.00s	0.37	3.77s	1.52
KT-5	3.21s	0.58	0.31	3.35s	0.54	0.69

Die negative Einschätzung des schwachen Schülers nach der experimentellen Bedingung 'Klassentest' setzt sich auf der Dimension S fort. HURRELMANN bezeichnet dies als "Vorstellung vom unangenehmen, dummen und ruppigen Rowdy"¹⁾. Der t-Wert ist signifikant: $t = 2.05s$. Zwischen den Versuchsgruppen KA-6 ($\bar{x} = 3.8$) und KT-6 ($\bar{x} = 3.5$) besteht keine signifikanter Unterschied, obwohl auch hier der Mittelwert der Gruppe, die den informellen Test bearbeiten mußte, niedriger ausfällt.

Das Bild des schwachen Schülers erscheint den Schülern des 5. Schuljahrs (KA-5: $\bar{x} = 4.5$) nach der Klassenarbeit auf der Dimension A 'gemächlich', 'ruhig' und 'still', während die Gruppe der älteren Schüler ihm nach der gleichen Bedingung (KA-6: $\bar{x} = 3.9$) die Eigenschaften 'eilig', 'lebhaft' und 'rege' prädiziert. Die Differenz der Mittelwerte ist signifikant: $t = 2.00s$.

1) HURRELMANN, K.: Unterrichtsorganisation und schulische Sozialisation. Eine Untersuchung zur Rolle der "Leistungs differenzierung" im schulischen Selektionsprozeß, Weinheim/Berlin/Basel 1971, S. 109

5.2.4 Einstufung des Konzepts 'GUTER SCHÜLER'

Die Mittelwerte aller Gruppen streuen auf der Dimension B über den positiven Bereich der Skala. Die experimentellen Bedingungen scheinen das Stereotyp 'guter Schüler' kaum zu affizieren; es besteht Einigkeit über seine Eigenschaften: Er ist 'fröhlich', 'freundlich', 'ehrlich', 'gut', 'klug' und 'angenehm'. Hatten die jüngeren Schüler der Gruppe KT-5 den schwachen Schüler sehr negativ bewertet, so ist es nun 'psycho'-logisch, daß der gute Schüler jetzt am höchsten eingestuft wird ($\bar{x}=6.0$). Die parallele Versuchsgruppe KA-5 schätzte ihn etwas niedriger ein ($\bar{x} = 5.7$). Die bei den älteren Schülern durch die experimentellen Bedingungen ausgelösten Diskriminationsprozesse verlaufen nicht so wie bei den Schülern des 5. Schuljahrs: KA-6 ($\bar{x} = 5.9$) und KT-6 ($\bar{x} = 5.6$). Signifikant unterscheiden sich die Gruppen KT-5 und KT-6: $t(KT-5/KT-6) = 2.31$.

Tabelle 24: Konzept XII: t-Werte der Dimensionen Bewertung, Stärke und Aktivität der Gruppen KA-5/KA-6, KA-5/KT-6, KT-5/KA-6 und KT-5/KT-6

	KA-6			KT-6		
	B	S	A	B	S	A
KA-5	0.77	-	1.19	0.62	1.16	1.26
KT-5	0.61	0.12	1.55	2.31s	0.94	1.63

Auch auf der Dimension S tendieren die älteren Schüler dazu, das Konzept 'guter Schüler' nach der Bedingung 'Klassentest' negativer einzustufen ($\bar{x} = 3.9$). Dem guten Schüler werden die Persönlichkeitseigenschaften 'hart', 'stark' und 'kraftvoll' zugeschrieben, so daß die Vorstellung vom Musterschüler ins Wanken gerät. Diese Korrektur des Stereotyps wird jedoch nicht von der Gruppe KA-6 vollzogen, deren 'implizite Persönlichkeitstheorie' die Einstufungen auf die Koordinaten 'weich', 'nachgiebig' und 'zart' richtet ($\bar{x} = 4.3$). Diese Urteilstendenz der Schüler des 6. Schuljahrs findet keine Entsprechung bei den vergleichbaren Gruppen des 5. Schuljahrs (KA-5: $\bar{x} = 4.3$, KT-5: $\bar{x} = 4.2$).

Die Mittelwerte der Gruppen KA-6 ($\bar{x} = 3.7$) und KT-6 ($\bar{x} = 3.7$) befinden sich auf der negativen Skalenseite der Dimension A. Mit den

Eigenschaften 'eilig', 'lebhaft' und 'rege' wird der gute Schüler charakterisiert. Wenn auch nicht signifikant, so doch der Tendenz nach, unterscheiden sich davon die positiven Einstufungen der Gruppen KA-5 ($\bar{x} = 4.1$) und KT-5 ($\bar{x} = 4.2$).

Eine interessante Hypothese, die auch pädagogisch-didaktische Relevanz besitzt, drängt sich als Nebenprodukt unserer bisherigen Auswertungen auf: Relativ starre Fixierungen, Stereotype, scheinen durch bestimmte Variablen veränderbar.

5.2.5. Einstufung des Konzepts 'ENGLISCHLEHRER'

Die Durchschnittswerte aller Gruppen rangieren im positiven Bereich der Dimension B, welche durch die Adjektive 'fröhlich', 'freundlich' usw. konstituiert wird. Kaum einen Unterschied ergeben die Mittelwerte der Gruppen KA-6 ($\bar{x} = 5.4$) und KT-6 ($\bar{x} = 5.3$). Auch die Differenz zwischen den Gruppen KA-5 ($\bar{x} = 4.6$) und KT-5 ($\bar{x} = 5.0$) liegt innerhalb zufälliger Variabilität. Trotz gleicher Versuchsbedingung unterscheiden sich die jüngeren Schüler (KA-5: $\bar{x} = 4.6$) von den älteren (KA-6: $\bar{x} = 5.4$) signifikant. Relativ weniger positiv wird das Konzept 'Englischlehrer' von den jüngeren Schülern eingestuft.

Tabelle 25: Konzept XIII: t-Werte der Dimensionen Bewertung, Stärke und Aktivität der Gruppen KA-5/KA-6, KA-5/KT-6, KT-5/KA-6 und KT-5/KT-6

	KA-6		A		KT-6	
	B	S	B	S	B	S
KA-5	2.73s	5.48s	0.67	1.20	3.15s	0.51
KT-5	1.46	5.86s	1.70	0.87	3.24s	0.40

Deutlich voneinander heben sich die im negativen Skalenbereich befindenden Mittelwerte der Gruppen KA-5 ($\bar{x} = 3.2$) und KT-5 ($\bar{x} = 3.3$) von den positiven Plazierungen der Gruppen KA-6 ($\bar{x} = 4.6$) und KT-6 ($\bar{x} = 3.3$) auf der Dimension S ab.

Halten die Schüler des 6. Schuljahrs den Englischlehrer für 'weich', 'nachgiebig' und 'zart', so sind die des 5. Jahrgangs auf einen 'harten', 'starken' und 'kraftvollen' Englischlehrer eingestellt. Interpretiert man den an der Grenze zur negativen Skalseite lokalisierten Mittelwert $\bar{x}(KT-6) = 4.1$ als negative Urteilstendenz

der Gruppe KT-6, die sich signifikant von der Gruppe KA-6 unterscheidet ($t=2.16$), so induziert auch hier die experimentelle Bedingung 'Klassentest' die wenig erfreuliche Identifikation mit dem Bild vom 'starken Mann', der durch die Eigenschaften 'hart', 'stark' und 'kraftvoll' charakterisiert wird.

Unsere in 5.2.4. aufgestellte ad-hoc-Hypothese müssen wir für die Deutung der im negativen Skalenbereich sich befindenden Mittelwerte der Versuchsgruppe '5. Schuljahr' dahingehend differenzieren, daß die Einstellungstereotype der jüngeren Schüler in unserem Versuch weniger von den experimentellen Bedingungen beeinflußt werden, weil ihre Wahrnehmungsfähigkeit für das soziale Objekt 'Lehrer' weniger ausgebildet ist.

Bei keiner Dimension der vorangehenden Konzepte zeigte sich eine so hohe Konformität der Urteile wie beim Konzept 'Engischlehrer': $\bar{x}(KA-6)=3.7$, $\bar{x}(KT-5)=3.3$, $\bar{x}(KA-5)=3.5$ und $\bar{x}(KT-6)=3.4$ auf der Dimension A. Übereinstimmend wird der Englischlehrer mit den Attributen 'eilig', 'lebhaft' und 'rege' versehen. Die Einstufungen auf dem Faktor Aktivität dieses Stereotyps scheinen durch die nur kurz einwirkenden experimentellen Bedingungen nicht beeinflußbar zu sein.¹⁾

5.2.6. Einstufungen der Konzepte 'KLASSENARBEIT', 'ZENSUR' und 'KLASSENTEST'

Wie die Tabellen 26, 27 und 28 zeigen, bestehen zwischen den Gruppen keine signifikanten Differenzen der Mittelwerte. Die Auswertungen in den folgenden Kapiteln werden jedoch bedeutsame Unterschiede ergeben, da dann der Grad der Ängstlichkeit berücksichtigt wird.²⁾

Tabelle 26: Konzept XIV: t-Werte der Dimensionen Bewertung, Stärke und Aktivität der Gruppen KA-5/KA-6, KA-5/KT-6, KT-5/KA-6 und KT-5/KT-6

	KA-6			KT-6		
	B	S	A	B	S	A
KA-5	1.01	0.54	0.18	1.35	0.04	0.96
KT-5	0.80	0.45	0.52	1.16	0.21	0.57

1) vgl. 4.5. (S. 50): Die Veränderbarkeit bestimmter Stereotype in Abhängigkeit von der Anzahl einstellungsbildender Erfahrungen bietet sich hier als Interpretationsschema an.

2) Dies hätte ein einfacher Versuchsplan nicht leisten können!

Tabelle 27: Konzept XV: t-Werte der Dimensionen Bewertung, Stärke und Aktivität der Gruppen KA-5/KA-6, KA-5/KT-6, KT-5/KA-6 und KT-5/KT-6

	KA-6			KT-6		
	B	S	A	B	S	A
KA-5	0.40	1.49	1.23	0.41	1.35	1.24
KT-5	0.52	0.37	0.46	0.39	0.11	0.42

Tabelle 28: Konzept XVI: t-Werte der Dimensionen Bewertung, Stärke und Aktivität der Gruppen KA-5/KA-6, KA-5/KT-6, KT-5/KA-6 und KT-5/KT-6

	KA-6			KT-6		
	B	S	A	B	S	A
KA-5	0.02	0.34	1.81	0.08	0.90	0.78
KT-5	0.17	0.09	1.77	0.87	0.45	0.58

5.3. Differentielle Analyse (1): Einstufungen der Konzepte IX - XVI durch wenig ängstliche und sehr ängstliche Schüler der Versuchsgruppe 5tes Schuljahr

Aus darstellungstechnischen Gründen werden wir die folgenden Unterkapitel nicht in der Reihenfolge der Hypothesen (4) - (10) anordnen, sondern nacheinander die Einstufungen der Konzepte IX - XVI auswerten, um auch die Hypothese (12) mit berücksichtigen zu können.

Für die statistische Auswertung benutzen wir die zweifache Varianzanalyse. Da wir uns nur für den Einfluß der Prüfinstrumente und der Ängstlichkeitsgrade interessieren, die im Experiment verwendet worden sind, liegt ein fixiertes Modell vor; d.h. wir stellen nur Schlußfolgerungen über die Unterschiede zwischen den Effekten der Bedingungen unseres Versuchsplanes an.

5.3.1. Konzept 'Selbstbild': Nicht die Art des Prüfinstruments, sondern der Grad der Ängstlichkeit determiniert das Selbstbild des Schülers auf den Dimensionen S und A

Mit der Hypothese (8) hatten wir auf der Dimension B einen Zusammenhang zwischen Ängstlichkeitsgrad und Art des Prüfinstruments postuliert: Sehr ängstliche und wenig ängstliche Schüler unter-

scheiden sich in ihrem Selbstbild im Anschluß an verschiedene Prüfungssituationen. Dies bedeutet in der statistischen Terminologie die Annahme einer Wechselwirkung. Der F_3 -Wert der Varianzanalyse ist jedoch nicht signifikant. Ebenfalls nur im Zufallsbereich liegen die Effekte der Prüfinstrumente. Wie die Mittelwerte der vier Gruppen zeigen: $\bar{x}(WÄ/IT)=5.3$, $\bar{x}(WÄ/KA)=5.2$, $\bar{x}(SÄ/IT)=5.1$ und $\bar{x}(SÄ/KA)=5.3$, halten sich die Schüler für 'fröhlich', 'freundlich' usw.

Tabelle 29: Tafel der Varianzanalyse: Konzept IX/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	70.20	99		
Zw. d. Prüfinstrumenten	0.11	1		$F_1 = 0.15$
Zw. WÄ und SÄ	0.04	1		$F_2 = 0.07$
Ww	0.63	1		$F_3 = 0.88$
Innerh. d. Z.	69.41	96	0.72	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Auf der Dimension S kann weder zwischen Ängstlichkeitsgrad und Prüfinstrument noch zwischen den beiden Prüfinstrumenten ein überzufälliger Zusammenhang nachgewiesen werden. Dagegen weist die statistische Analyse einen sehr signifikanten F-Wert zwischen wenig und sehr ängstlichen Kindern auf ($F_2=12.21$). Wenig ängstliche Schüler wählen die Adjektive 'hart', 'stark' und 'kraftvoll', sehr ängstliche 'weich', 'nachgiebig' und 'zart': $\bar{x}(WÄ/IT,KA)=3.5$. und $\bar{x}(SÄ/IT,KA)=4.4$.

Tabelle 30: Tafel der Varianzanalyse: Konzept IX/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	254.46	99		
Zw. d. Prüfinstrumenten	0.49	1		$F_1 = 0.35$
Zw. WÄ und SÄ	17.33	1		$F_2 = 12.21ss$
Ww	0.40	1		$F_3 = 0.29$
Innerh. d. Z.	136.24	96	1.42	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Die Einstufungen zum Konzept 'Selbstbild' verlaufen, was die signifikanten F-Werte betrifft, auf der Dimension A fast parallel zur

Dimension S. Wenig ängstliche Schüler ($\bar{x}(\text{WÄ/IT,KA})=3.4$) ordnen ihrem Selbstbild die Attribute 'eilig', 'lebhaft' und 'rege' zu, sehr ängstliche ($\bar{x}(\text{SÄ/IT,KA})=4.2$) dagegen die Merkmale 'gemächlich', 'ruhig' und 'still'.

Tabelle 31: Tafel der Varianzanalyse: Konzept IX/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	168.25	99		
Zw. d. Prüfinstrumenten	0.12	1		$F_1 = 0.08$
Zw. WÄ und SÄ	15.09	1		$F_2 = 9.64_{ss}$
Ww	2.73	1		$F_3 = 1.74$
Innerh. d. Z.	150.31	96	1.57	

$$F_{0.05;1/96}=3.96 \quad F_{0.01;1/96}=6.91$$

5.3.2. Konzept 'WUNSCHBILD': Der informelle Test bewirkt Dissoziation:

Die Diskrepanz zwischen 'Selbstbild' und 'Wunschbild' wird vergrößert

Die Mittelwerte aller Gruppen sind auf der positiven Skalenseite der Dimension B lokalisiert, doch bescheinigen sich die sehr und wenig ängstlichen Schüler, die dem Einfluß der experimentellen Bedingung 'Klassentest' ausgesetzt worden sind, ein sehr signifikantes Mehr an Freundlichkeit, Fröhlichkeit usw. als die Schüler, welche auf die Bedingung 'Klassenarbeit' reagierten ($F_1=6.99_{ss}$). Vielleicht

Tabelle 32: Tafel der Varianzanalyse: Konzept X/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	75.06	99		
Zw. d. Prüfinstrumenten	5.01	1		$F_1 = 6.99_{ss}$
Zw. WÄ und SÄ	0.03	1		$F_2 = 0.04$
Ww	1.15	1		$F_3 = 1.60$
Innerh. d. Z.	68.88	96	0.72	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

kann dies als Gegenreaktion auf die Frustrationen auslösende Druck-situation des informellen Tests interpretiert werden. Eingeschränkt

wird diese Hypothese allerdings dadurch, daß sich wenig und sehr ängstliche Schüler in ihren Einstufungen nicht signifikant unterscheiden.

Keine signifikanten F-Werte brachte die Varianzanalyse auf der Dimension S. Es fällt jedoch der hohe Mittelwert der Gruppe SÄ/KA auf ($\bar{x}(SÄ/KA)=4.6$), dem der fast signifikante F_3 -Wert zu verdanken ist. Die restlichen drei Mittelwerte befinden sich auf der negativen Skalenseite ($\bar{x}(WÄ/IT)=3.7$, $\bar{x}(WÄ/KA)=3.4$ und $\bar{x}(SÄ/IT)=3.6$).

Tabelle 33: Tafel der Varianzanalyse: Konzept X/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	287.28	99		
Zw. d. Prüf-instrumenten	2.54	1		$F_1 = 0.92$
Zw. WÄ und SÄ	8.02	1		$F_2 = 2.89$
Ww	10.43	1		$F_3 = 3.76$
Innerh. d. Z.	266.29	96	2.77	

$$F_{0.05;1/90}=3.14 \quad F_{0.01;1/96}=6.91$$

Da es folgeschwer wäre, die Nullhypothese abzulehnen, nehmen wir an, daß der Tendenz nach eine Wechselwirkung besteht. Die Mittelwertdifferenzen wurden deshalb mit dem DUNCAN-Test auf Signifikanzen untersucht. Signifikant erwies sich dabei die Mittelwertdifferenz von wenig und sehr ängstlichen Schülern, welche die Klassenarbeit geschrieben hatten. Es scheint demnach folgende Wechselwirkung zu bestehen: Sehr ängstliche Schüler stufen sich nach der Klassenarbeit auf dem Konzept 'Wunschbild' mit den Adjektiven 'weich', 'nachgiebig' und 'zart', wenig ängstliche nach derselben Bedingung 'hart', 'stark' und 'kraftvoll' ein.

Nach den Bedingungen 'Klassenarbeit' und 'Klassentest' stufen sich auf der Dimension A alle Gruppen ($\bar{x}(WÄ/IT)=3.7$, $\bar{x}(WÄ/KA)=3.9$, $\bar{x}(SÄ/IT)=3.9$ und $\bar{x}(SÄ/KA)=4.0$) in der Nähe der neutralen Skalenmitte ein. Den geringen Mittelwertdifferenzen entsprechen die nicht signifikanten niedrigen F-Werte der Varianzanalyse.

Tabelle 34: Tafel der Varianzanalyse: Konzept X/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	263.20	99		
Zw. d. Prüf- instrumenten	0.43	1		$F_1 = 0.16$
Zw. WÄ und SÄ	0.72	1		$F_2 = 0.26$
Ww	0.02	1		$F_3 = 0.01$
Innerh. d. Z.	262.03	96	2.73	
$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$				

5.3.3. Konzept 'SCHWACHER SCHÜLER': Der informelle Test hat einen dissoziativen Sozialisationseffekt: Der leistungsschwache Schüler wird nicht integriert

Wenn wir weiter oben davon ausgegangen waren, daß Prüfinstrumente Wirkungen auf schulische Sozialisationsprozesse ausüben, so müßten besonders deutliche Effekte bei der Einstufung des Konzepts 'schwacher Schüler' auftreten, denn nach der Analyse der Syntax informeller Tests postulierten wir Dissoziationen bei dieser Art von Leistungsüberprüfung. Hypothetisch konnten wir annehmen, daß das Stereotyp des schwachen Schülers nach dem Klassentest in den Vorstellungen der Schüler negativere Züge tragen müßte als nach der Klassenarbeit.

Die statistische Auswertung der Einstufungen auf der Dimension B offenbarte ein sich sehr signifikant unterscheidendes Urteilsverhalten ($F_1=10.28_{ss}$). Wenig ängstliche und sehr ängstliche Schüler stimmen nach dem informellen Test darin überein, daß der leistungsschwache Schüler 'traurig', 'unfreundlich', 'unehrlich', 'schlecht', 'dumm' und 'unangenehm' ist ($\bar{x}(WÄ,SÄ/IT)=3.3$). Der Mittelwert der sehr ängstlichen Schüler ($\bar{x}(SÄ/IT)=3.0$) indiziert dabei eine noch negativere Bewertungsrichtung als der Mittelwert der Gruppe der wenig ängstlichen Schüler ($\bar{x}(WÄ/IT)=3.6$). Anders verhalten sich bei ihren Einstufungen die Gruppen der sehr ängstlichen und wenig ängstlichen Kinder nach der herkömmlichen Klassenarbeit ($\bar{x}(WÄ/KA)=4.2$, $\bar{x}(SÄ/KA)=4.1$). Sie billigen ihrem leistungsschwachen Mitschüler noch etwas Fröhlichkeit und Freundlichkeit zu.

Tabelle 35: Tafel der Varianzanalyse: Konzept XI/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	198.78	99		
Zw. d. Prüf-instrumenten	18.87	1		$F_1 = 10.28_{ss}$
Zw. WÄ und SÄ	1.10	1		$F_2 = 0.60$
Ww.	2.63	1		$F_3 = 1.43$
Innerh. d. Z.	176.19	96	1.84	
$F_{0.05;1/96} = 3.94 \quad F_{0.01;1/96} = 6.91$				

Das gleiche Urteilsverhalten setzt sich auf der Dimension S fort. Wenig ängstliche und sehr ängstliche Schüler erleben nach der Test-situation ($\bar{x}(WÄ/IT)=3.7$, $\bar{x}(SÄ/IT)=3.5$) den schwachen Schüler als 'hart', 'stark' und 'kraftvoll'. Die Bedingung 'Klassenarbeit' induziert ein durch die Adjektive 'weich', 'nachgiebig' und 'zart' gekennzeichnetes Feld ($\bar{x}(WA/KA)=4.4$, $\bar{x}(SA/KA)=4.5$). Das Urteilsverhalten unterscheidet sich sehr signifikant ($F_1=8.01_{ss}$).

Tabelle 36: Tafel der Varianzanalyse: Konzept XI/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	229.97	99		
Zw. d. Prüf-instrumenten	17.67	1		$F_1 = 8.01_{ss}$
Zw. WA und SA	0.07	1		$F_2 = 0.03$
Ww.	0.45	1		$F_3 = 0.20$
Innerh. d. Z.	211.77	96	2.21	
$F_{0.05;1/96} = 3.94 \quad F_{0.01;1/96} = 6.91$				

Auch auf der Dimension A finden wir den gleichen Zusammenhang: Die sich unterscheidenden Prüfungsarten rufen ein sich signifikant unterscheidendes Urteilsverhalten hervor ($F_1=4.40_{ss}$). Es zeigt sich

Tabelle 37: Tafel der Varianzanalyse: Konzept XI/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	250.44	99		
Zw. d. Prüf-instrumenten	10.91	1		$F_1 = 4.40_{ss}$
Zw. WÄ und SÄ	1.00	1		$F_2 = 0.40$
Ww.	0.28	1		$F_3 = 0.11$
Innerh. d. Z.	238.24	96	2.48	
$F_{0.05;1/96} = 3.94 \quad F_{0.01;1/96} = 6.91$				

die gleiche Richtung der Einstufungen wie auf den beiden anderen Dimensionen. Mit den Eigenschaften 'eilig', 'lebhaft' und 'rege' wird der schwache Schüler nach dem Klassentest versehen ($\bar{x}(\text{WÄ, SÄ/IT})=3.8$), während ihm nach der Klassenarbeit Merkmale zugeschrieben werden, die sonst dem guten, braven und leistungsstarken Schüler zuerteilt werden: Gemächlichkeit, ruhiges und stilles Benehmen.

5.3.4. Konzept 'GUTER SCHÜLER': Der Tendenz nach bewerten sehr ängstliche Schüler das Konzept auf der Dimension B positiver als wenig ängstliche

Das Konzept XII bezieht sich auf Schüler mit guter Leistungsposition. "Schüler mit guter Leistungsposition werden ... von ihren Mitschülern in Bezug auf sozial unerwünschte Eigenschaften und Verhaltensmerkmale weit positiver eingestuft als andere. Es handelt sich dabei um solche Merkmale, die weit mehr als das schulische Rollenverhalten der Kinder betreffen und sich auf ganz allgemeine Interaktionsqualitäten erstrecken (z.B. "freundlich - unfreundlich", "angenehm - unangenehm")."¹⁾

Weiter oben äußerten wir die Vermutung, daß das Einstufungsverhalten bei bestimmten Konzepten wenig veränderbar zu sein scheint. Dies ist der Fall beim Konzept 'guter Schüler', denn alle Mittelwerte befinden sich dicht beieinander auf der oberen Hälfte der positiven Skalenseite der Dimension B ($\bar{x}(\text{WÄ/IT})=5.9$, $\bar{x}(\text{WÄ/KA})=5.6$, $\bar{x}(\text{SÄ/IT})=6.0$ und $\bar{x}(\text{SÄ/KA})=6.0$ und differieren nur wenig. Signifikanzen bestehen nicht.

Tabelle 38: Tafel der Varianzanalyse: Konzept XII/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	106.71	99		
Zw. den Prüfinstrumenten	1.67	1		$F_1=1.60$
Zw. WÄ und SÄ	2.74	1		$F_2=2.63$
Ww.	2.08	1		$F_3=2.00$
Innerh. d. Z.	100.22	96	1.04	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Nicht mehr so ausgeprägt erscheint das Stereotyp des guten Schülers auf der Dimension S. Weniger ängstliche Schüler ($\bar{x}(\text{WÄ/IT})=4.1$ und $\bar{x}(\text{WÄ/KA})=4.0$) überschreiten mit ihren Einstufungen fast die Grenze

zum negativen Skalenbereich, während die sehr ängstlichen Schüler diesem Trend nicht folgen ($\bar{x}(\text{SÄ/IT})=4.4$, $\bar{x}(\text{SÄ/KA})=4.4$). Dieser Unterschied ist aber statistisch nicht bedeutsam. Die Prüfungsbedingungen rufen ebenfalls keine signifikanten Effekte hervor.

Tabelle 39: Tafel der Varianzanalyse: Konzept XII/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	280.30	99		
Zw. den Prüf-instrumenten	0.05	1		$F_1 = 0.02$
Zw. WÄ und SÄ	3.07	1		$F_2 = 1.06$
Ww.	0.05	1		$F_3 = 0.02$
Innerh. d. Z.	277.14	96	2.89	
$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$				

Ebenfalls nicht signifikant sind die Einstufungen auf der Dimension A.

Tabelle 40: Tafel der Varianzanalyse: Konzept XII/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	286.38	99		
Zw. d. Prüf-instrumenten	0.58	1		$F_1 = 0.20$
Zw. WÄ und SÄ	1.22	1		$F_2 = 0.41$
Ww.	0.70	1		$F_3 = 0.24$
Innerh. d. Z.	283.89	96	2.96	
$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$				

5.3.5. Konzept 'ENGLISCHLEHRER': Weder die Prüfinstrumente noch der Grad der Ängstlichkeit determinieren signifikant die Einstufungen

Die Hypothese zum Konzept 'Englischlehrer' lautete für die Dimension B: Aufgrund der experimentellen Bedingung 'Klassentest' wird das Konzept 'Englischlehrer' von sehr ängstlichen Schülern negativ bewertet. Die Argumentationsfigur, die zur Bildung dieser Hypothese führte, ist folgende: Instruktion und Bearbeitungszeiten sind durch das Tonband festgelegt; der Zeitdruck wirkt frustrierend und löst Aggressionen aus, die auf die Einstufungen der Konzepte durchschlagen. Wie jedoch die Tabelle 41 ausweist, ist diese Hypo-

these nicht zu halten; auch wenn man das Signifikanzniveau erhöht, um einen Fehler zweiter Art zu vermeiden. Denn in unserem Fall ist es ja folgenscher, die Nullhypothese fälschlicherweise anzunehmen. Der Tendenz nach scheinen sich die experimentellen Bedingungen unterschiedlich auszuwirken ($\bar{x}(W\ddot{A}/IT)=5.1$, $\bar{x}(S\ddot{A}/IT)=4.9$, $\bar{x}(W\ddot{A}/KA)=4.5$ und $\bar{x}(S\ddot{A}/KA)=4.6$).

-Tabelle 41: Tafel der Varianzanalyse: Konzept XIII/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	242.37	99		
Zw. d. Prüf-instrumenten	4.47	1		$F_1 = 1.81$
Zw. W\ddot{A} und S\ddot{A}	0.00	1		
Ww.	0.73	1		$F_3 = 0.30$
Innerh. d. Z.	237.16	96	2.47	

$$F_{0.05;1/96} = 3.94 \quad F_{0.01;1/96} = 6.91$$

Die Vorstellung vom Englischlehrer rückt auf der Dimension S in die Nähe des schwachen Schülers mit rowdyhaftem Betragen, der sich in Konfliktsituationen mit physischer Kraft durchsetzt. Sowohl die wenig ängstlichen ($\bar{x}(W\ddot{A}/IT)=3.3$, $\bar{x}(W\ddot{A}/KA)=3.1$) als auch die sehr ängstlichen Schüler ($\bar{x}(S\ddot{A}/IT)=3.3$, $\bar{x}(S\ddot{A}/KA)=3.4$) gehen konform in ihrem Urteil, daß der Englischlehrer 'hart', 'stark' und 'kraftvoll' ist. Er erscheint ihnen als 'starker Mann'.

Tabelle 42: Tafel der Varianzanalyse: Konzept XIII/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	166.78	99		
Zw. d. Prüf-instrumenten	0.08	1		$F_1 = 0.05$
Zw. W\ddot{A} und S\ddot{A}	0.42	1		$F_2 = 0.25$
Ww.	0.51	1		$F_3 = 0.30$
Innerh. d. Z.	165.77	96	1.73	

$$F_{0.05;1/96} = 3.94 \quad F_{0.01;1/96} = 6.91$$

Wenden wir uns nun den Mittelwerten auf der Dimension A zu ($\bar{x}(W\ddot{A}/IT)=3.5$, $\bar{x}(W\ddot{A}/KA)=3.6$, $\bar{x}(S\ddot{A}/IT)=3.1$ und $\bar{x}(S\ddot{A}/KA)=3.3$), so ist nach der Tabelle 43 keine Mittelwertsdifferenz statistisch bedeutsam. Nur der F_2 -Wert ($F_2=2.06$) deutet die statistisch nicht gesicherte Tendenz an, daß wenig ängstliche Schüler und sehr ängst-

liche sich in ihrem Einstufungsverhalten unterscheiden. Beide Gruppen sprechen dem Englischlehrer die Eigenschaften 'eilig', 'lebhaft' und 'rege' zu.

Tabelle 43: Tafel der Varianzanalyse: Konzept XIII/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	215.15	99		
Zw. den Prüf-instrumenten	1.37	1		$F_1 = 0.33$
Zw. WÄ und SÄ	4.49	1		$F_2 = 2.06$
Ww.	0.04	1		$F_3 = 0.02$
Innerh. d. Z.	209.25	96	2.18	
$F_{0.05;1/96} = 3.94 \quad F_{0.01;1/96} = 6.91$				

5.3.6. Konzept 'KLASSENARBEIT': Wenig ängstliche und sehr ängstliche zeigen nach der Bedingung 'informeller Test' auf der Dimension B sich signifikant unterscheidende Einstufungen

Unsere Hypothese (5) hieß: Ein Prüfinstrument, das weniger 'autoritär' anmutet, wird von sehr ängstlichen Schülern positiv beurteilt. Vergleichen wir den Mittelwert der sehr ängstlichen Schüler nach der Bedingung 'Klassentest' ($\bar{x}(SA/KT)=4.2$) mit dem der sehr ängstlichen nach der Bedingung 'Klassentest' ($\bar{x}(SA/KA)=4.3$), so erscheint unsere Vermutung bestätigt. Dafür spricht auch der signifikante F-Wert der Wechselwirkung ($F_3=4.12s$). Um nun auszumachen, ob die Mittelwertsdifferenz dieser beiden Gruppen signifikant ist, prüften wir mit dem DUNCAN-Test. Der Unterschied ist jedoch kleiner als der vom DUNCAN-Test geforderte ($0.6 < 0.8$), so daß die Nullhypothese beibehalten werden mußte. Da aber das Beharren auf der Nullhypothese folgen-schwer ist, prüften wir mit niedriger Verlässlichkeit. Mit dieser Einschränkung können wir behaupten, daß zwischen Ängstlichkeitsgrad und Art der Leistungsüberprüfung eine Wechselwirkung besteht: Sehr ängstliche Schüler beurteilen das Konzept 'Klassenarbeit' positiver, wenn sie eine Klassenarbeit geschrieben haben, als wenn sie einen Test bearbeiten müßten. Wenig ängstliche Kinder stufen auch nach der Bedingung 'informeller Test' das Konzept 'Klassenarbeit' hoch ein ($\bar{x}(WÄ/KT)=5.0$, $\bar{x}(SÄ/KT)=4.2$). Sie unterscheiden sich signifikant von ihren sehr ängstlichen Mitschülern.

Tabelle 44: Tafel der Varianzanalyse: Konzept XIV/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	172.97	99		
Zw. d. Prüf-instrumenten	0.11	1		$F_1 = 0.06$
Zw. WÄ und SÄ	1.76	1		$F_2 = 1.03$
Ww.	7.04	1		$F_3 = 4.12s$
Innerh. d. Z.	164.06	96	1.71	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=9.91$$

Die Tabelle 45 enthält für die Dimension S keinen einzigen signifi- kanten F-Wert. Das Konzept 'Klassenarbeit' wird von allen Gruppen ($\bar{x}(WÄ/KT)=3.8$, $\bar{x}(WÄ/KA)=3.8$, $\bar{x}(SÄ/KT)=3.7$ und $\bar{x}(SÄ/KA)=3.6$) negativ erlebt. Beide Arten der Leistungsüberprüfung werden als 'hart', 'stark' und 'kraftvoll' empfunden.

Tabelle 45: Tafel der Varianzanalyse: Konzept XIV/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	176.24	99		
Zw. d. Prüf-instrumenten	0.03	1		$F_1 = 0.02$
Zw. WÄ und SÄ	0.56	1		$F_2 = 0.31$
Ww.	0.18	1		$F_3 = 0.10$
Innerh. d. Z.	175.47	96	1.83	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Auf der Dimension A beurteilen wenig ängstliche Kinder die Klassenarbeit mit den Eigenschaften 'gemächlich', 'ruhig' und 'still' ($\bar{x}(WÄ/KT)=4.1$, $\bar{x}(WÄ/KA)=4.4$), sehr ängstliche dagegen mit 'eilig', 'lebhaft' und 'rege' ($\bar{x}(SÄ/KT)=3.9$, $\bar{x}(SÄ/KA)=3.7$). Da sich der F_2 -Wert relativ nahe an der erforderlichen Prüfgröße befindet und es

Tabelle 46: Tafel der Varianzanalyse: Konzept XIV/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	165.78	99		
Zw. d. Prüf-instrumenten	0.18	1		$F_1 = 0.11$
Zw. WÄ und SÄ	5.09	1		$F_2 = 3.07$
Ww.	1.17	1		$F_3 = 0.71$
Innerh. d. Z.	159.35	96	1.66	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

folgeschwer wäre, einen Fehler zweiter Art zu begehen, fassen wir den soeben beschriebenen Zusammenhang als Tendenz auf. Eine endgültige Klärung bleibt weiteren Untersuchungen überlassen.

5.3.7. Konzept 'ZENSUR': Das Stereotyp 'Zensur' wird weder durch die Prüfungsart noch durch den Ängstlichkeitsgrad verändert

Das Experiment war kurz vor den Versetzungszeugnissen durchgeführt worden, so daß anzunehmen war, daß Zensuren eine große Rolle im Erleben der Kinder spielten. Entgegen unseren Erwartungen verteilen sich die Einstufungen auf der positiven Skalenseite der Dimension B. Die wenig ängstlichen ($\bar{x}(W\ddot{A}/KT)=5.0$, $\bar{x}(W\ddot{A}/KA)=4.8$) und sehr ängstlichen Schüler ($\bar{x}(S\ddot{A}/KT)=4.8$, $\bar{x}(S\ddot{A}/KA)=4.8$) sind in ihrem Urteilsverhalten konform. Die Interiorisationsprozesse, die zur Bildung des positiv besetzten Stereotyps 'Zensur' geführt haben, scheinen abgeschlossen und so verfestigt zu sein, daß momentane Einstellungsänderungen durch die Bedingungen 'Klassenarbeit' oder 'informeller Test' nicht ausgelöst werden können.

Tabelle 47: Tafel der Varianzanalyse: Konzept XV/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	166.96	99		
Zw. d. Prüf-instrumenten	0.00	1		F_1
Zw. WÄ und SÄ	1.04	1		$F_2 = 0.60$
Ww.	0.96	1		$F_3 = 0.56$
Innerh. d. Z.	164.99	96	1.72	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Ein weniger eindeutiges Bild bietet sich bei den Einstufungen auf der Dimension S, obwohl die Mittelwerte nur zufällig voneinander abweichen. Nach dem informellen Test ordnen wenig ängstliche und sehr ängstliche Schüler ($\bar{x}(W\ddot{A}/KT)=3.9$, $\bar{x}(W\ddot{A}/KA)=3.7$) dem Konzept 'Zensur' die Attribute 'hart', 'stark' und 'kraftvoll' zu, während nach der Klassenarbeit die Urteile an den Anfang des positiven Skalenbereichs plaziert werden ($\bar{x}(S\ddot{A}/KT)=4.1$, $\bar{x}(S\ddot{A}/KA)=4.0$). Das Konzept 'Zensur' mutet diese Schüler eher 'weich', 'nachgiebig' und 'zart' an.

Tabelle 48: Tafel der Varianzanalyse: Konzept XV/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	174.03	99		
Zw. d. Prüf-instrumenten	2.21	1		$F_1 = 1.24$
Zw. WÄ und SÄ	0.97	1		$F_2 = 0.55$
Ww.	0.20	1		$F_3 = 0.11$
Innerh. d. Z.	170.65	96	1.78	

$$F_{0.05;1/96} = 3.94 \quad F_{0.01;1/96} = 6.91$$

Wie schon beim Faktor 'Stärke' zeigen sich auch zwischen den Mittelwerten der Dimension A ($\bar{x}(WÄ/KT)=4.1$, $\bar{x}(WÄ/KA)=4.2$, $\bar{x}(SÄ/KT)=3.9$ und $\bar{x}(SÄ/KA)=4.1$) keine statistisch bedeutsamen Differenzen. Das Schwanken der Mittelwerte um die neutrale Skalenmitte findet vielleicht sein psychisches Korrelat in der Haltung, daß Zensuren etwas Endgültiges darstellen, die Urteilsabgabe somit gleichgültig erfolgt.

Tabelle 49: Tafel der Varianzanalyse: Konzept XV/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	135.05	99		
Zw. d. Prüf-instrumenten	0.66	1		$F_1 = 0.47$
Zw. WÄ und SÄ	1.11	1		$F_2 = 0.80$
Ww.	0.08	1		$F_3 = 0.06$
Innerh. d. Z.	133.20	96	1.39	

$$F_{0.05;1/96} = 3.94 \quad F_{0.01;1/96} = 6.91$$

5.3.8. Konzept 'KLASSENTEST': Sehr ängstliche Schüler stufen den informellen Test auf der Dimension S signifikant positiver ein als wenig ängstliche

Das Konzept 'Klassentest' wird von allen Versuchspersonen positiv eingestuft ($\bar{x}(WÄ/KT)=5.0$, $\bar{x}(WÄ/KA)=4.7$, $\bar{x}(SÄ/KT)=4.9$ und $\bar{x}(SÄ/KA)=4.8$). Kein F-Wert ist signifikant. Zieht man zum Vergleich die Mittelwerte des Konzepts 'Klassenarbeit' heran, so zeigt sich, daß die Schüler des 5. Schuljahrs auf der Bewertungsdimension zwischen beiden Konzepten nicht diskriminieren ($\bar{x}=4.7$).

Tabelle 50: Tafel der Varianzanalyse: Konzept XVI/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	184.57	99		
Zw. d. Prüfinstrumenten	0.65	1		$F_1 = 0.34$
Zw. WÄ und SÄ	0.00	1		$F_2 = 0.00$
Ww.	1.13	1		$F_3 = 0.59$
Innerh. d. Z.	182.79	96	1.90	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Wie aus der Tabelle 51 zu ersehen ist, bewirken die Prüfinstrumente keine statistisch gesicherten Unterschiede beider abhängigen Variablen. Überzufällig dagegen unterscheiden sich in ihrem Urteilsverhalten die wenig ängstlichen ($\bar{x}(WÄ/KT)=3.7$, $\bar{x}(WÄ/KA)=3.6$) von den sehr ängstlichen Schülern ($\bar{x}(SÄ/KT)=4.1$, $\bar{x}(SÄ/KA)=4.5$) auf der Dimension S. Ein geringer Ängstlichkeitsgrad verursacht Einstufungen des Konzepts 'Klassentest', die mit den Adjektiven 'hart', 'stark' und 'kraftvoll' beschrieben werden können; sehr ängstliche Kinder des 5. Jahrgangs bevorzugen die Merkmale 'weich', 'nachgiebig' und 'zart'.

Tabelle 51: Tafel der Varianzanalyse: Konzept XVI/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	175.91	99		
Zw. d. Prüfinstrumenten	0.25	1		$F_1 = 0.15$
Zw. WÄ und SÄ	10.85	1		$F_2 = 6.36s$
Ww.	1.07	1		$F_3 = 0.62$
Innerh. d. Z.	163.75	96	1.71	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Tabelle 52: Tafel der Varianzanalyse: Konzept XVI/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	168.47	99		
Zw. d. Prüfinstrumenten	0.09	1		$F_1 = 0.05$
Zw. WÄ und SÄ	0.31	1		$F_2 = 0.18$
Ww.	1.17	1		$F_3 = 0.67$
Innerh. d. Z.	166.90	96	1.74	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Die Mittelwerte der Einstufungen auf der Dimension A ($\bar{x}(W\ddot{A}/KT)=4.1$, $\bar{x}(W\ddot{A}/KA)=4.0$, $\bar{x}(S\ddot{A}/KT)=4.0$ und $\bar{x}(S\ddot{A}/KA)=4.3$) gruppieren sich auf der positiven Skalenseite in der Nähe der neutralen Mitte.

5.3.9. Zusammenfassung der Ergebnisse

Als theoretisches Modell, das die Genese der Prüfungsangst "nicht nur aus intrapersonalen Komponenten der individuellen Sozialisation allein erklärt, sondern auch aus den situativen Komponenten der spezifischen Prüfungssituation"¹⁾, zogen wir die habit-interference-Theorie von MANDLER und SARASON heran. Nach dieser Prüfungsangst-Theorie werden aufgabenirrelevante Verhaltensweisen gelernt, die in jeder neuen Prüfungssituation die Leistungsfähigkeit des Schülers mindern, die sich "durch das häufige Auftreten von Angstsituationen allmählich verselbständigen, nicht mehr an prüfungsähnliche Situationen gebunden sind und sozusagen als überdauernde Determinanten des Verhaltens auch ohne den Auslöser 'Lehrer' oder 'Prüfung' das Individuum in seiner Leistungsfähigkeit hemmen."²⁾

Die Institution Schule als Abbild gesellschaftlicher Bedingungen, so hatten wir angenommen, induziert Angst aufgrund permanenter Leistungsforderungen und produziert aufgabenirrelevantes Verhalten. Da Prüfungen in einer hochdifferenzierten Industriegesellschaft wahrscheinlich notwendig sind und auch nicht abgeschafft werden können, sollte die Prüfungssituation zumindest so gestaltet werden, daß sie ein Mindestmaß an Dissoziativität aufweist.

Im Mittelpunkt unseres Forschungsinteresses stand dabei nicht die Frage nach dem Zusammenhang von Angst und Leistung (Schulleistung), sondern das Interesse an den Nebenwirkungen normorientierter Leistungsmessung: Wie sieht der Schüler sich selbst nach einer hochstrukturierten Aufgabendarbietung, wie nach einer Klassenarbeit? Zu dieser Problemstellung fanden wir im deutschsprachigen Raum nur eine Untersuchung, die zum Teil unsere Problemstellung thematisierte³⁾, jedoch nur an PH-Studenten durchgeführt worden ist.

1) PRELL, S.: Die Auswirkung von Prüfungsangst auf verschiedene Formen der Lernzielkontrolle, in: Psychologie in Erziehung und Unterricht, 1973, S. 17

2) SCHELL, H.: Angst und Schulleistung, Göttingen 1972, S. 22

3) PRELL, S., a.a.O., S. 15

1. Weiter oben stellten wir die Hypothese auf: Sehr ängstliche und wenig ängstliche Schüler unterscheiden sich im Anschluß an verschiedene Prüfungssituationen in ihrem Selbstbild. Diese Vermutung wurde nicht bestätigt. Es zeigte sich, daß das Selbstbild sehr signifikant auf den Dimensionen S und A durch den Ängstlichkeitsgrad determiniert wird, nicht jedoch auf der Bewertungsdimension. Wenig ängstliche Schüler wählen auf der Dimension S die Eigenschaften 'hart', 'stark' und 'kraftvoll', sehr ängstliche 'weich', 'nachgiebig' und 'zart'. Die Präferenz sehr ängstlicher Schüler für die zuletzt genannten Adjektive steht repräsentativ für geringeres Durchsetzungsvermögen und Zurückweichen vor Aufgabenforderungen, da "die Einstellung zur eigenen Person ... beeinträchtigt"¹⁾ ist. Das gleiche Verhalten erscheint auch bei den Einstufungen auf der Dimension A. Mit 'eilig', 'lebhaft' und 'rege' bezeichnen sich wenig ängstliche Schüler, mit 'gemächlich', 'ruhig' und 'still' sehr ängstliche.

2. Klassenarbeit und informeller Test bewirken auf der Dimension B des Konzepts 'Wunschbild' sich sehr signifikant unterscheidende Einstufungen im positiven Bereich der Skala. Das Mehr an positiver Bewertung des Konzepts nach der Bedingung 'informeller Test' führen wir auf den Frustrationen auslösenden Zeitdruck dieser Prüfungsart zurück, der den Wunsch nach mehr Freundlichkeit und Fröhlichkeit hervorruft. Eine Wechselwirkung kann auf der Dimension S angenommen werden: Sehr ängstliche Schüler wünschen nach einer Klassenarbeit 'weich', 'nachgiebig' und 'zart' zu sein, wenig ängstliche dagegen 'hart', 'stark' und 'kraftvoll'. Diese Eigenschaften werden auch von den übrigen zwei Gruppen, die den Test bearbeiten mußten, gewählt. Die Bedingung 'informeller Test' ruft damit den Wunsch nach Verhaltensweisen hervor, die, negativ gedeutet, dem Rowdy-Verhalten nahekommen und somit unerwünschte Sozialisationseffekte auslösen.

Wenden wir uns nun dem Vergleich der Mittelwerte von Selbstbild und Wunschbild zu, da wir die Hypothese aufgestellt hatten: Wunschbild und Selbstbild stimmen nach hochstrukturierten Prüfungssituationen mehr überein als nach wenig strukturierten. Wie jedoch die Tabelle 53

1) NICKEL, H./SCHLÜTER, P./FENNER, H.-J.: Angstwerte, Intelligenztest- und Schulleistungen sowie der Einfluß der Lehrerpersönlichkeit bei Schülern verschiedener Schularten, in: Psychologie in Erziehung und Unterricht, 1973, S. 1

zeigt, wird unsere Annahme nicht bestätigt. Die Differenzen zwischen Selbstbild und Wunschbild sind nach dem informellen Test bei allen Gruppen und auf allen Dimensionen größer als nach der Klassenarbeit. Zwischen Wunsch- und Selbstbild besteht nach der Bedingung 'informeller Test' eine größere Diskrepanz als nach der Klassenarbeit. Diese Diskrepanz interpretieren wir als dissoziatives Moment.

Tabelle 53: Mittelwerte der Konzepte IX - XVI auf den Dimensionen Bewertung, Stärke und Aktivität

	WÄ/KT	SÄ/KT	WÄ/KA	SÄ/KA
B: IX	5.3	5.1	5.2	5.3
X	6.1	6.4	5.9	5.7
S: IX	3.4	4.1	3.5	4.4
X	3.7	3.6	3.4	4.6
A: IX	3.1	4.2	3.5	3.9
X	3.7	3.9	3.9	4.0

3. Zum Stereotyp 'schwacher Schüler' liegen Untersuchungen vor, in denen er z.B. mit den Attributen 'faul', 'bequem' und 'träge' charakterisiert wird. Unser Merkmalsraum wird aus den Eigenschaften 'traurig', 'hart', 'unfreundlich', 'dumm' usw. gebildet. Man könnte also auch bei unseren Versuchsgruppen so rigide, fixierte Einstellungen erwarten, daß ein Prüfungsinstrument keine Effekte mehr hervorrufen kann. Doch auf allen Dimensionen wird das Urteilsverhalten der Schüler durch die Art der Leistungsüberprüfung determiniert. Nach der Bedingung 'informeller Test' assoziieren die Schüler mit dem Bild des schwachen Schülers auf der Dimension B die Attribute 'traurig', 'unfreundlich', 'unehrlich', 'schlecht', 'dumm' und 'unangenehm', auf der Dimension S 'hart', 'stark' und 'kraftvoll', auf der Dimension A 'eilig', 'rege' und 'lebhaft'. Denselben Eindruck haben auch sehr ängstliche Schüler nach der Klassenarbeit. Von allen anderen Gruppen, die die Klassenarbeit schreiben mußten, wird das Bild des schwachen Schülers auf allen Dimensionen auf die positive Skalenseite projiziert. Diese unterschiedlichen Urteilsreaktionen sind auf den Dimensionen B und S sehr signifikant. Pointiert formuliert, erscheint die Klassenarbeit so etwas wie einen Solidarisierungseffekt auszulösen, während der informelle Test das Bild des 'Aggressors' entstehen läßt, d.h. der leistungsschwächere Schüler

wird nicht in die Klasse integriert, sondern zum Außenseiter degradiert. Wieweit dieses Untersuchungsergebnis nur momentane Einstellungsänderungen reflektiert, müßte durch einen Versuchsplan mit Zeitreihen überprüft werden.

4. Das Stereotyp 'guter Schüler' scheint so fixiert zu sein, daß die kurz andauernde Prüfungssituation ohne Wirkung bleibt. Der relativ hohe F-Wert auf der Dimension B ($F=2.63$), zeigt die Tendenz sehr ängstlicher Schüler, ihre leistungsstarken Mitschüler höher einzustufen, als es wenig ängstlichere tun.

5. In Bezug auf das Konzept 'Engischlehrer' hatten wir die Hypothese gebildet, das Konzept werde von sehr ängstlichen Schülern negativ beurteilt. Sie konnte nicht bestätigt werden. Weder der Grad der Ängstlichkeit noch die Art der Prüfung bewirkten auf irgendeiner Dimension einen bedeutsamen Unterschied.

6. Es wurde vermutet, daß ein Prüfinstrument, das die Schüler weniger autoritär anmutet, von sehr ängstlichen Schülern positiv bewertet würde. Die Einstufungstendenz ist fast signifikant, d.h. sehr ängstliche Schüler beurteilen das Prüfungsinstrument 'Klassenarbeit', mit dem ihre Leistung gemessen worden ist, positiver als Schüler des gleichen Persönlichkeitsmerkmals nach der Testarbeit. Statistisch gesichert ist der Unterschied zwischen wenig ängstlichen und sehr ängstlichen Schülern, denen der informelle Test administriert worden ist. Wenig ängstliche Schüler billigen dem informellen Test mehr 'Freundlichkeit', 'Fröhlichkeit' usw. zu als sehr ängstliche. Auf der Dimension Aktivität scheint das Einstufungsverhalten der Tendenz nach eine Funktion des Ängstlichkeitsgrades zu sein: Wenig ängstliche Schüler empfinden die Klassenarbeit noch als 'ruhig', 'still' und 'gemächlich', während ihre Antipoden das gleiche Prüfinstrument als 'eilig', 'lebhaft' und 'rege' einstufen. Die pädagogische Konsequenz wäre, einen Kompromiß in der Form der Leistungsüberprüfung zu finden, in dem unterschiedliche Ängstlichkeitsgrade berücksichtigt werden. Außerdem müßte untersucht werden, ob dieses Problem auch bei lernzielorientierter Leistungsmessung auftritt.

7. Das Konzept 'Zensur' wird auf der Dimension B von allen Schülern positiv eingestuft. Normkonformität charakterisiert ihr verbales

Verhalten. Nicht so eindeutig in die positive Skalenrichtung tendieren die Urteile auf den Dimensionen S und A. Wenig und sehr ängstliche Schüler wählen nach der Bedingung 'Klassentest' die Attribute 'hart', 'stark' und 'kraftvoll', nach der Bedingung 'Klassenarbeit' jedoch 'weich', 'nachgiebig' und 'zart'. Allerdings sind diese Unterschiede nicht signifikant. Auf der Dimension A streuen die Urteile um die neutrale Skalenmitte.

8. Die Schüler des 5. Schuljahrs diskriminieren nicht zwischen den Konzepten 'Klassenarbeit' und 'Klassentest'. Für beide Konzepte werden auf der Dimension B positive Urteile abgegeben. Die Einstufungen auf dem Faktor Bewertung scheinen eher durch die Übernahme allgemein akzeptierter Normen determiniert zu werden als durch die real erlebte Situation des Geprüftwerdens. Die Einstufungen auf den Dimensionen S und A scheinen valider für die Einschätzung der eigenen Situation. Signifikant unterscheiden sich wenig ängstliche Schüler von sehr ängstlichen auf der Dimension S. Ein geringer Ängstlichkeitsgrad lokalisiert das Konzept 'Klassentest' auf die mit den Adjektiven 'hart', 'stark' und 'kraftvoll' definierte Skalseite. Sehr ängstliche Schüler des 5. Jahrgangs neigen dazu, das gleiche Konzept 'weich', 'nachgiebig' und 'zart' einzuschätzen. Auf der Dimension A streuen die Mittelwerte aller Gruppen um die neutrale Skalenmitte.

5.4. Differentielle Analyse (2): Einstufungen der Konzepte IX - XVI durch wenig ängstliche und sehr ängstliche Schüler der Versuchsgruppe 6tes Schuljahr

Wenn auch der Allgemeinheitsgrad der Differenzierungshypothese zu groß ist, um für inhaltlich fest umrissene Problemstellungen spezielle Prognosen zu deduzieren, so kann doch die Hypothese niederen Allgemeinheitsgrades angenommen werden: Schüler des 6. Jahrgangs besitzen eine gegliedertere Wahrnehmungsfähigkeit als ihre jüngeren Schulkameraden, so daß bei den Einstufungen der älteren Kinder mit mehr Wechselwirkungen und Effekten der Prüfinstrumente zu rechnen ist.

5.4.1. Konzept 'Selbstbild': Das Selbstbild sehr ängstlicher Schüler wird auf der Dimension S durch das Prüfinstrument determiniert

Wie im 5. Schuljahr stufen sich alle vier Versuchsgruppen auf der

positiven Skalenseite der Dimension B ein ($\bar{x}(W\ddot{A}/KT)=5.4$ (5.3)¹⁾, $\bar{x}(W\ddot{A}/KA)=5.6$ (5.2), $\bar{x}(S\ddot{A}/KT)=5.2$ (5.1) und $\bar{x}(S\ddot{A}/KA)=5.3$ (5.3), d.h. sie schätzen sich als 'fröhlich', 'freundlich' usw. ein. Die Prüfungsbedingungen können ihr Realbild nicht verändern.

Tabelle 54: Tafel der Varianzanalyse: Konzept IX/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	68.32	99		
Zw. d. Prüf-instrumenten	0.56	1		$F_1 = 0.82$
Zw. W $\ddot{A}$ und S $\ddot{A}$	1.49	1		$F_2 = 2.17$
Ww.	0.42	1		$F_3 = 0.61$
Innerh. d. Z.	65.85	96	0.68	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Ausgenommen die Gruppe S $\ddot{A}$ /KT ($\bar{x}(S\ddot{A}/KT)=3.8$ (4.1), versehen alle anderen Gruppen ihr Selbstbild auf der Dimension S mit einem Pluszeichen ($\bar{x}(W\ddot{A}/KT)=4.4$ (3.4), $\bar{x}(W\ddot{A}/KA)=4.2$ (3.5), $\bar{x}(S\ddot{A}/KA)=4.6$ (4.4). Im 5. Schuljahr hatten wir einen signifikanten Unterschied zwischen wenig ängstlichen und sehr ängstlichen Schülern gefunden. Bei den älteren Schülern deutet der fast signifikante F_3 -Wert auf eine Wechselwirkung hin. Das ist ein Indiz dafür, daß sie ein differenzierteres Urteilsverhalten aufweisen. Da es uns zu folgens schwer erscheint, die Nullhypothese anzunehmen, interpretieren wir den F_3 -Wert als Tendenz. Unter dieser Voraussetzung wandten wir den DUNCAN-Test zur Identifizierung der bedeutsamen Mittelwertsdifferenzen an. Fast signifikant erweist sich dabei die Differenz der Mittelwerte $\bar{x}(S\ddot{A}/KA)$ und $\bar{x}(S\ddot{A}/KT)$: 0.800 - 0.862. Sehr ängstliche Schüler ordnen nach der Leistungsprüfung mit einem informellen Test ihrem Selbstbild die Eigenschaften 'hart', 'stark' und 'kraftvoll' zu, während die Gruppe mit dem gleichen Ängstlichkeitsgrad ihr Realbild nach der Klassenarbeit mit den Adjektiven 'weich', 'nachgiebig' und 'zart' metaphorisch umschreibt.

Alle Einstufungen auf der Dimension A befinden sich im negativen Bereich der Skala ($\bar{x}(W\ddot{A}/KT)=2,6$ (3.1), $\bar{x}(W\ddot{A}/KA)=3.7$ (3.5), $\bar{x}(S\ddot{A}/KT)=3.4$ (4.2) und $\bar{x}(S\ddot{A}/KA)=3.6$ (3.9)). Der Unterschied zwischen den Prüfinstrumenten ist nicht signifikant. ($F_1=2.12$).

¹⁾ Mittelwerte des 5. Schuljahrs in Klammern

Grafik 4: Arithmetische Mittel der Dimensionen Bewertung, Stärke und Aktivität der Konzepte IX - XVI unter den experimentellen Bedingungen 'Klassemarbeit' und 'Klassentest' bei wenig ängstlichen und sehr ängstlichen Schülern des 6. Schuljahrs

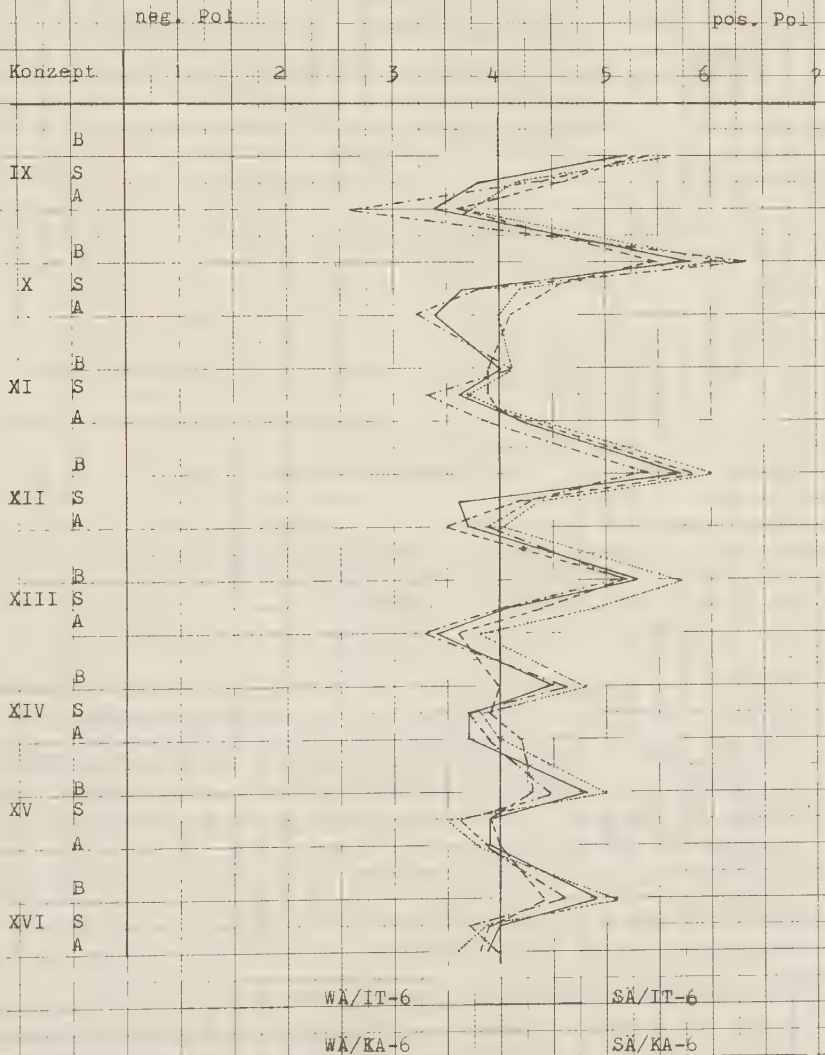


Tabelle 55: Tafel der Varianzanalyse: Konzept IX/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	199.15	99		
Zw. d. Prüf-instrumenten	2.05	1		$F_1 = 1.04$
Zw. WÄ und SÄ	0.19	1		$F_2 = 0.09$
Ww.	6.58	1		$F_3 = 3.32$
Innerh. d. Z.	190.33	96	1.98	
$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$				

Tabelle 56: Tafel der Varianzanalyse: Konzept IX/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	478.15	99		
Zw. d. Prüf-instrumenten	10.14	1		$F_1 = 2.12$
Zw. WÄ und SÄ	2.77	1		$F_2 = 0.58$
Ww.	5.85	1		$F_3 = 1.22$
Innerh. d. Z.	459.39	96	4.72	
$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$				

5.4.2. Konzept "WUNSCHBILD": Der informelle Test bewirkt Dissoziation: Persönlichkeitseigenschaften werden bevorzugt, die aggressives, konkurrentes Verhalten begünstigen

Wie auf der Dimension B zu erwarten ist, siedeln alle Gruppen ihre Wunschvorstellungen vom idealen Ich im positiven Bereich der Skala an. Der relativ hohe F_1 -Wert weist der Tendenz nach auf unterschiedliche Effekte der beiden Prüfinstrumente hin. Im 5. Schuljahr waren die durch die Prüfungsbedingungen hervorgerufenen Differenzen der Mittelwerte sehr signifikant ($F_1=6.99ss$). Auf dem 1%-Niveau signifikant ist der Unterschied zwischen wenig ängstlichen und sehr ängstlichen Schülern ($F_2=9.62ss$): $\bar{x}(WÄ/KT)=6.3$ (6.1), $\bar{x}(WÄ/KA)=6.1$ (5.9), $\bar{x}(SÄ/KT)=5.8$ (6.4) und $\bar{x}(SÄ/KT)=5.5$ (5.7).

Die Einstufungen auf der Dimension S werden durch die Prüfungsinstrumente determiniert. Wenig ängstliche und sehr ängstliche Schüler wünschen sich nach dem informellen Test die Eigenschaften 'hart', 'stark' und 'kraftvoll' ($\bar{x}(WÄ/KT)=3.8$ (3.7), $\bar{x}(SÄ/KT)=3.7$ (3.6)),

während sich die vergleichbare Gruppe nach der Klassenarbeit für die Adjektive 'weich', 'nachgiebig' und 'zart' entscheidet:

$\bar{x}(\text{WÄ/KA})=4.2$ (3.4), $\bar{x}(\text{SÄ/KA})=4.6$ (4.6).

Tabelle 57: Tafel der Varianzanalyse: Konzept X/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	77.29	99		
Zw. d. Prüfinstrumenten	2.00	1		$F_1 = 2.82$
Zw. WÄ und SÄ	6.84	1		$F_2 = 9.62_{ss}$
Ww.	0.23	1		$F_3 = 0.33$
Innerh. d. Z.	68.22	96	0.71	

$$F_{0.05;1/96}=3. + F_{0.01;1/96}=6.91$$

Tabelle 58: Tafel der Varianzanalyse: Konzept X/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	223.30	99		
Zw. d. Prüfinstrumenten	11.56	1		$F_1 = 5.29_{ss}$
Zw. WÄ und SÄ	0.37	1		$F_2 = 0.17$
Ww.	1.60	1		$F_3 = 0.73$
Innerh. d. Z.	209.77	96	2.17	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Die Bedingung 'informeller Test' ruft demnach den Wunsch nach Persönlichkeitsmerkmalen hervor, die konkurrentes Verhalten begünstigen.

Ausgeprägter als bei den Versuchsgruppen des 6ten Schuljahrs fallen auch die Einstufungen auf der Dimension A aus. Zwischen beiden Prüfinstrumenten besteht ein sehr signifikanter Unterschied: $F_1=8.45_{ss}$.

Tabelle 59: Tafel der Varianzanalyse: Konzept X/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	189.00	99		
Zw. d. Prüfinstrumenten	15.21	1		$F_1 = 8.45_{ss}$
Zw. WÄ und SÄ	1.05	1		$F_2 = 0.58$
Ww.	0.01	1		$F_3 = 0.01$
Innerh. d. Z.	172.73	96	1.80	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Die sehr ängstlichen und wenig ängstlichen Schüler, die der Bedingung 'Klassentest' ausgesetzt worden waren, identifizieren sich mit den Eigenschaften 'eilig', 'lebhaft' und 'rege' ($\bar{x}(\text{WÄ/KT})=3.2$ (3.7), $\bar{x}(\text{SÄ/KT})=3.4$ (3.9), $\bar{x}(\text{WÄ/KA})=4.0$ (3.9) und $\bar{x}(\text{SÄ/KA})=4.1$ (4.0)), während die beiden Gruppen, die die Klassenarbeit geschrieben haben, ihr Wunschbild mit den Adjektiven 'weich', 'nachgiebig' und 'zart' kennzeichnen.

5.4.5. Konzept 'SCHWACHER SCHÜLER': Die Einstufungen auf den Dimensionen B, S und A werden weder durch den Grad der Ängstlichkeit noch durch die Art des Prüfinstrumente determiniert

Im Gegensatz zu den Daten der Versuchsgruppe 5tes Schuljahr stehen die im 6ten Schuljahr erhobenen. Waren dort die Einstufungen des Konzepts 'schwacher Schüler' auf allen Dimensionen eine Funktion der Prüfinstrumente, so streuen hier die Mittelwerte über alle Gruppen hinweg um die neutrale Skalenmitte ($\bar{x}(\text{WÄ/IT})=4.1$ (3.6), $\bar{x}(\text{WÄ/KA})=4.1$ (4.1), $\bar{x}(\text{SÄ/KT})=4.0$ (3.0) und $\bar{x}(\text{SÄ/KA})=3.9$ (3.6).

Tabelle 60: Tafel der Varianzanalyse: Konzept XI/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	88.10	99		
Zw. d. Prüfinstrumenten	0.09	1		$F_1 = 0.10$
Zw. WÄ und SÄ	0.25	1		$F_2 = 0.27$
Ww.	0.25	1		$F_3 = 0.27$
Innerh. d. Z.	87.50	96	0.93	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Auf der Dimension S schlägt bei allen Gruppen ($\bar{x}(\text{WÄ/KT})=3.3$ (3.7), $\bar{x}(\text{WÄ/KA})=3.7$ (4.4), $\bar{x}(\text{SÄ/KT})=3.6$ (3.5) und $\bar{x}(\text{SÄ/KA})=3.9$ (4.5)) das Stereotyp des leistungsschwachen Schülers durch. Man prädiiziert ihm die Eigenschaften 'hart', 'stark' und 'kraftvoll'. Wir bezeichneten das bildhaft als Rowdy-Syndrom. Das soziale Objekt 'schwacher Schüler' scheint in den Vorstellungen der Versuchsschüler des 6. Schuljahrs das fixierte Ergebnis schulischer Sozialisationsprozesse, so daß auch eine momentane Einstellungsänderung durch die Prüfinstrumente nicht erzielt werden kann.

Tabelle 61: Tafel der Varianzanalyse: Konzept XI/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	147.61	99		
Zw. d. Prüfinstrumenten	2.46	1		$F_1 = 1.65$
Zw. WÄ und SÄ	2.08	1		$F_2 = 1.39$
Ww.	0.09	1		$F_3 = 0.06$
Innerh. d. Z.	142.99	96	1.49	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Auf der Dimension A streuen die Mittelwerte aller Gruppen mit relativ kleinen Differenzen um die neutrale Mitte der Skala ($\bar{x}(WÄ/KT)=3.9$ (3.9), $\bar{x}(WÄ/KA)=3.0$ (4.5), $\bar{x}(SÄ/KT)=4.2$ (3.7) und $\bar{x}(SÄ/KA)=4.0$ (4.4)). Eine Stereotypisierung des Konzepts ist nicht erkennbar.

Tabelle 62: Tafel der Varianzanalyse: Konzept XI/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	182.81	99		
Zw. d. Prüfinstrumenten	0.36	1		$F_1 = 0.19$
Zw. WÄ und SÄ	0.36	1		$F_2 = 0.19$
Ww.	1.29	1		$F_3 = 0.69$
Innerh. d. Z.	180.79	96	1.88	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

5.4.4. Konzept 'GUTER SCHÜLER': Die Prüfbedingungen lösen auf der Dimension B der Tendenz nach unterschiedliche Effekte aus

Weniger fixiert als bei den Schülern des 5. Schuljahrs scheint das Stereotyp 'guter Schüler' in den Vorstellungen der älteren Kinder zu sein. Dies zeigen deutlich die Mittelwerte der Versuchsgruppen: $\bar{x}(WÄ/KT)=5.4$ (5.9), $\bar{x}(WÄ/KA)=5.9$ (5.6), $\bar{x}(SÄ/KT)=5.7$ (6.0) und $\bar{x}(SÄ/KA)=5.8$ (6.0). Der relativ hohe - jedoch nicht signifikante - F_1 -Wert indiziert die Tendenz, daß die Prüfinstrumente unterschiedliche Effekte erzeugen. Die älteren Kinder scheinen das Konzept 'guter Schüler' nach der Klassenarbeit durchschnittlich höher einzustufen als nach dem informellen Test. Wenig ängstliche Schüler weisen dabei den größten Einstellungsunterschied auf.

Tabelle 63: Tafel der Varianzanalyse: Konzept XII/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	70.89	99		
Zw. d. Prüf-instrumenten	2.11	1		$F_1 = 3.01$
Zw. WÄ und SÄ	0.07	1		$F_2 = 0.10$
Ww.	1.39	1		$F_3 = 1.99$
Innerh. d. Z.	67.32	96	0.70	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Die Tendenz, im guten Schüler einen Musterschüler zu sehen, der 'weich', 'nachgiebig' und 'zart' ist, kehrt bei drei Gruppen des 6. Jahrgangs auf der Dimension S wieder ($\bar{x}(WÄ/KT)=4.3$ (4.1), $\bar{x}(WÄ/KA)=4.4$ (4.0) und $\bar{x}(SÄ/KA)=4.2$ (4.4)). Die Gruppe der sehr ängstlichen Schüler, denen der informelle Test gegeben worden ist ($\bar{x}(SÄ/KT)=3.6$ (4.4)), lokalisiert ihre Urteile jedoch auf der negativen Skalenseite. Dem leistungsstarken Schüler werden die Eigenschaften 'hart', 'stark' und 'kraftvoll' zugeschrieben. Wenn sich auch das Urteilsverhalten dieser Versuchsgruppe deskriptiv und qualitativ von den anderen Schülergruppen abhebt, so ist zu bedenken, daß die statistische Analyse ($F_2=2.58$) keine Signifikanz aufwies. Dennoch billigen wir der qualitativen Analyse heuristischen Wert zu. Wir deuten den F_2 -Wert als Tendenz.

Tabelle 64: Tafel der Varianzanalyse: Konzept XII/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	181.01	99		
Zw. d. Prüf-instrumenten	2.84	1		$F_1 = 1.58$
Zw. WÄ und SÄ	4.61	1		$F_2 = 2.58$
Ww.	1.56	1		$F_3 = 0.87$
Innerh. d. Z.	171.99	96	1.79	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Im Gegensatz zu den Schülern des 5. Schuljahrs stufen im 6. Schuljahr alle vier Gruppen das Konzept XII auf der Dimension A im negativen Skalenbereich ein ($\bar{x}(WÄ/KT)=3.9$ (4.4), $\bar{x}(WÄ/KA)=3.9$ (4.3), $\bar{x}(SÄ/KT)=3.7$ (4.0) und $\bar{x}(SÄ/KA)=3.5$ (4.0)). Strebt das durchschnitt-

liche Urteilsverhalten der wenig ängstlichen Schüler auf die neutrale Skalenmitte zu, so scheint sich das der sehr ängstlichen von ihr zu entfernen.

Tabelle 65: Tafel der Varianzanalyse: Konzept XII/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	169.22	99		
Zw. d. Prüf- instrumenten	0.19	1		$F_1 = 0.11$
Zw. WÄ und SÄ	2.25	1		$F_2 = 1.30$
Ww.	0.93	1		$F_3 = 0.54$
Innerh. d. Z.	165.85	96	1.73	

$$F_{0.05;1/96}=3.4 \quad F_{0.01;1/96}=6.91$$

5.4.5. Konzept 'ENGLISCHLEHRER': Die Bedingung 'informeller Test' bewirkt auf der Dimension S weniger positive Einstufungen als die Bedingung 'Klassenarbeit'

Die Schüler des 6. Schuljahrs stufen das Konzept 'Englischlehrer' relativ höher ein als die des 5. ($\bar{x}(WÄ/KT)=5.2$ (5.1), $\bar{x}(WÄ/KA)=5.6$ (4.9), $\bar{x}(SÄ/KT)=5.3$ (4.5) und $\bar{x}(SÄ/KA)=5.2$ (4.6)). Der Englischlehrer erscheint ihnen 'freundlicher', 'fröhlicher', 'angenehmer' usw. Das Bild des Englischlehrers ähnelt auf der Dimension B sehr der Vorstellung vom guten Schüler. Zwischen den Gruppen bestehen keine signifikanten Unterschiede. Das Urteilsverhalten der Schüler ist so starr, daß die unterschiedlichen Prüfungsbedingungen das Stereotyp auf der Bewertungsdimension nicht verändern können.

Tabelle 66: Tafel der Varianzanalyse: Konzept XIII/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	144.67	99		
Zw. d. Prüf- instrumenten	0.64	1		$F_1 = 0.44$
Zw. WÄ und SÄ	1.36	1		$F_2 = 0.93$
Ww.	1.70	1		$F_3 = 1.16$
Innerh. d. Z.	140.96	96	1.47	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Alle Versuchsgruppen des 5. Schuljahrs sprachen dem Konzept 'Englischlehrer' auf der Dimension S die Eigenschaften 'hart', 'stark'

und 'kraftvoll' zu. Von den Versuchsgruppen des 6. Schuljahrs scheinen nur die in die Richtung des negativen Pols zu tendieren, denen der informelle Test administriert worden ist ($\bar{x}(\text{WÄ/KT})=4.1$ (3.3) und $\bar{x}(\text{SÄ/KT})=4.1$ (3.3)). Sie heben sich signifikant von den restlichen zwei Gruppen ($\bar{x}(\text{WÄ/KA})=4.9$ (3.1) und $\bar{x}(\text{SÄ/KA})=4.4$ (3.8) ab ($F_1=4.88s$).

Tabelle 67: Tafel der Varianzanalyse: Konzept XII/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	159.65	99		
Zw. d. Prüfinstrumenten	7.13	1		$F_1 = 4.58s$
Zw. WÄ und SÄ	1.62	1		$F_2 = 1.04$
Ww.	1.61	1		$F_3 = 1.03$
Innerh. d. Z.	149.29	96	1.56	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Auf der Dimension A zeigt sich der Tendenz nach ein unterschiedlicher Effekt der Prüfinstrumente ($F_1=2.17ns$). 'Eilig', 'lebhaft' und 'rege' ist der Englischlehrer nach dem Urteil aller Gruppen ($\bar{x}(\text{WÄ/KT})=3.3$ (3.5), $\bar{x}(\text{WÄ/KA})=3.8$ (3.6), $\bar{x}(\text{SÄ/KT})=3.4$ (3.1) und $\bar{x}(\text{SÄ/KA})=3.6$ (3.3)). Die Übereinstimmung mit den Urteilen der Schüler aus dem 7. Schuljahr ist deutlich.

Tabelle 68: Tafel der Varianzanalyse: Konzept XIII/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	115.68	99		
Zw. d. Prüfinstrumenten	2.54	1		$F_1 = 2.17$
Zw. WÄ und SÄ	0.22	1		$F_2 = 0.19$
Ww.	0.25	1		$F_3 = 0.21$
Innerh. d. Z.	112.67	96	1.17	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

5.4.6. Konzept 'KLASSENARBEIT': Nach der Klassenarbeit beurteilen sehr ängstliche Schüler im Gegensatz zu wenig ängstlichen das Konzept 'Klassenarbeit' auf der Dimension B signifikant negativer

Deutlich positiv sind die Einstufungen auf der Dimension B bei drei Gruppen ($\bar{x}(\text{WÄ/KT})=4.6$ (5.0), $\bar{x}(\text{WÄ/KA})=4.8$ (4.6) und $\bar{x}(\text{SÄ/KT})=4.5$ (4.2)).

Neutral verhält sich die Gruppe ängstlicher Schüler, welche die Klassenarbeit schreiben mußte ($\bar{x}(SA/KA)=4.0$ (4.8)). Dieses auffällige Urteilsverhalten korrespondiert dem signifikanten F_3 -Wert, der eine Wechselwirkung anzeigt. Wir prüften alle Mittelwertsdifferenzen mit dem DUNCAN-Test auf Signifikanz. Die Analyse brachte ein Überzu-

Tabelle 69: Tafel der Varianzanalyse: Konzept XIV/B-6

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	111.37	99		
Zw. d. Prüf-instrumenten	0.16	1		$F_1 = 0.14$
Zw. WÄ und SÄ	1.97	1		$F_2 = 1.82$
Ww.	5.62	1		$F_3 = 5.208$
Innerh. d. Z.	103.63	96	1.08	

$$F_{0.05;1/96} = 3.96 \quad F_{0.01;1/96} = 6.91$$

fällig sich unterscheidendes Urteilsverhalten zwischen wenig ängstlichen und sehr ängstlichen Schülern, die der Bedingung 'Klassenarbeit' ausgesetzt worden waren. Sehr ängstliche Schüler stehen nach dem Schreiben einer Klassenarbeit dieser Tätigkeit weniger positiv gegenüber als ihre weniger ängstlichen Schulkameraden. Eine ähnliche Bewertung der Klassenarbeit fanden wir auch bei den sehr ängstlichen Kindern des 5. Schuljahrs, die den informellen Test bearbeitet hatten. Sie bewerteten die Klassenarbeit signifikant weniger positiv als ihre weniger ängstlichen Mitschüler. Auch dort tendierte der Mittelwert der sehr ängstlichen Schüler zur neutralen Mitte der Skala.

Auf der Dimension S differieren die Einstufungen der vier Gruppen nur innerhalb des Zufallsbereichs. Gemeinsam ist allen durchschnittlichen Einstufungen, daß sie sich im negativen Skalenbereich befinden:

Tabelle 70: Tafel der Varianzanalyse: Konzept XIV/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	92.42	99		
Zw. d. Prüf-instrumenten	0.53	1		$F_1 = 0.56$
Zw. WÄ und SÄ	0.00	1		$F_2 = 0.00$
Ww.	0.02	1		$F_3 = 0.02$
Innerh. d. Z.	91.87	96	0.96	

$$F_{0.05;1/96} = 3.96 \quad F_{0.01;1/96} = 6.91$$

$\bar{x}(W\ddot{A}/KT)=3.7$ (3.8), $\bar{x}(W\ddot{A}/KA)=3.8$ (3.8), $\bar{x}(S\ddot{A}/KT)=3.7$ (3.7) und $\bar{x}(S\ddot{A}/KA)=3.9$ (3.6)). Die Präferenz der Schüler für die Anmutungsqualitäten 'hart', 'stark' und 'kraftvoll' entspricht den druckauslösenden Momenten der Situation.

Die quantitative statistische Analyse weist auf der Dimension A nur zufällige Unterschiede zwischen den Mittelwerten auf ($\bar{x}(W\ddot{A}/KT)=3.9$ (4.1), $\bar{x}(W\ddot{A}/KA)=4.0$ (4.4), $\bar{x}(S\ddot{A}/KT)=3.7$ (3.9) und $\bar{x}(S\ddot{A}/KA)=4.2$ (3.7)). Wenig ängstliche und sehr ängstliche Schüler stufen die Klassenarbeit nach dem Test mit den Adjektiven 'eilig', 'lebhaft' und 'rege' ein. Die Bedingung 'Klassenarbeit' verlagert die Urteile, wenn auch nicht signifikant auf die Koordinaten 'gemächlich', 'ruhig' und 'still' des Faktors Aktivität.

Tabelle 71: Tafel der Varianzanalyse: Konzept XIV/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	94.25	99		
Zw. d. Prüf-instrumenten	1.68	1		$F_1 = 1.76$
Zw. WÄ und SÄ	0.00	1		$F_2 = 0.00$
Ww.	1.06	1		$F_3 = 1.11$
Innerh. d. Z.	91.51	96	0.95	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=0.91$$

5.4.7. Konzept 'ZENSUR': Nach der Bedingung 'Klassenarbeit' beurteilen sehr ängstliche Schüler das Konzept auf der Dimension B negativer als wenig ängstliche

Die Schüler des 6. Schuljahrs urteilen differenzierter als die des 5. Jahrgangs. Hatten die jüngeren Schüler über alle Gruppen hinweg das Konzept 'Zensur' auf der Dimension B positiv bewertet, so tendiert die Gruppe der sehr ängstlichen älteren Schüler mit ihren Einstufungen zur neutralen Skalenmitte ($\bar{x}(S\ddot{A}/KA)=4.2$ (4.8)). Von diesem Mittelwert heben sich die durchschnittlichen Einschätzungen der drei anderen Gruppen ab ($\bar{x}(W\ddot{A}/KT)=4.5$ (5.0), $\bar{x}(W\ddot{A}/KA)=5.0$ (4.8) und $\bar{x}(S\ddot{A}/KT)=4.8$ (4.6)). Es besteht eine signifikante Wechselwirkung ($F_3=4.448$). Der DUNCAN-Test ergab einen fast signifikanten Unterschied zwischen den Gruppen SÄ/KA und WÄ/KA ($0.745 < 0.751$). Die Nullhypothese wurde dennoch nicht angenommen, da ein Fehler zweiter Art durch-

aus möglich ist. Sieht man von den Urteilsreaktionen der Gruppen SÄ/KA ab, so ist es erstaunlich, daß der Überwiegende Teil der Schüler das Konzept 'Zensur' als 'fröhlich', 'gut', 'angenehm' usw. empfindet.

Tabelle 72: Tafel der Varianzanalyse: Konzept XV/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	153.82	99		
Zw. d. Prüf-instrumenten	0.02	1		$F_1 = 0.01$
Zw. WÄ und SÄ	1.27	1		$F_2 = 0.84$
Ww.	6.75	1		$F_3 = 4.44s$
Innerh. d. Z.	145.78	96	1.52	

$$F_{0.05;1/96} = 3.94 \quad F_{0.01;1/96} = 6.91$$

Betrachtet man dagegen das Urteilsverhalten der Schüler auf der Dimension S, so stehen diese Einstufungen im Widerspruch zur Bewertung des Konzepts. Wenn man das Konzept 'Zensur' mit den Eigenschaften 'fröhlich', 'freundlich' usw. qualifiziert, dann müßten konsequenterweise die Adjektive 'weich', 'nachgiebig' und 'zart' gewählt werden, nicht jedoch die Merkmale 'hart', 'stark' und 'kraftvoll'. Nur die sehr ängstlichen Schüler ($\bar{x}(SÄ/KT)=3.9$ (3.7) und $\bar{x}(SÄ/KA)=3.9$ (4.1)) vermeiden der Tendenz nach diesen Widerspruch. Ihre Reaktionen unterscheiden sich fast signifikant von denen der wenig ängstlichen ($\bar{x}(WÄ/KT)=3.6$ (3.9) und $\bar{x}(WÄ/KA)=3.4$ (4.0): $F_2=3.38ns$.

Tabelle 73: Tafel der Varianzanalyse: Konzept XV/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	113.98	99		
Zw. d. Prüf-instrumenten	0.13	1		$F_1 = 0.12$
Zw. WÄ und SÄ	3.86	1		$F_2 = 3.38$
Ww.	0.13	1		$F_3 = 0.12$
Innerh. d. Z.	109.85	96	1.14	

$$F_{0.05;1/96} = 3.94 \quad F_{0.01;1/96} = 6.91$$

Die Einstufungen auf der Dimension A nähern sich der neutralen Skalenmitte: $\bar{x}(WÄ/KT)=3.9$ (4.1), $\bar{x}(WÄ/KA)=3.8$ (4.2), $\bar{x}(SÄ/KT)=3.9$ (3.9) und $\bar{x}(SÄ/KA)=4.0$ (4.1)). Die Schüler des 5. und 6. Schuljahrs

gehen fast konform mit ihren Einstufungen. Vergleicht man die durchschnittlichen Einstufungen auf allen Dimensionen, so manifestiert sich ein ambivalentes Urteilsverhalten, das vielleicht als Widerspruch zwischen formellen und informellen Normen interpretiert werden könnte.

Tabelle 74: Tafel der Varianzanalyse: Konzept XIV/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	80.11	99		
Zw. d. Prüf-instrumenten	0.01	1		$F_1 = 0.01$
Zw. WÄ und SÄ	0.22	1		$F_2 = 0.26$
Ww.	0.28	1		$F_3 = 0.34$
Innerh. d. Z.	79.61	96	0.83	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

5.4.8. Konzept 'KLASSENTEST': Wenig ängstliche Schüler bewerten den informellen Test nach der Bedingung 'Klassenarbeit' positiver als sehr ängstliche

Auch das Konzept 'Klassentest' wird von den Schülern des 6. Schuljahrs auf der Dimension B differenzierter eingestuft als von denen des 5. Jahrgangs ($\bar{x}(S\bar{A}/KT)=4.6$ (5.0), $\bar{x}(W\bar{A}/KA)=5.1$ (4.7), $\bar{x}(S\bar{A}/KT)=4.9$ (4.9) und $\bar{x}(S\bar{A}/KA)=4.4$ (4.8)), wenn man den signifikanten F-Wert der Wechselwirkung als Kriterium akzeptiert. Wie schon beim Konzept 'Klassenarbeit' unterscheiden sich auch hier die Gruppen WÄ/KA und SÄ/KA bedeutsam. Die Differenz dieser Mittelwerte ist größer als die mit dem DUNCAN-Test ermittelte Prüfgröße: $0.773 > 0.703$. Wenig ängstliche Schüler beurteilen nach der Klassenarbeit das Konzept 'Klassentest' positiver als sehr ängstliche.

Tabelle 75: Tafel der Varianzanalyse: Konzept XVI/B

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	135.87	99		
Zw. d. Prüf-instrumenten	0.21	1		$F_1 = 0.15$
Zw. WÄ und SÄ	1.34	1		$F_2 = 1.02$
Ww.	7.13	1		$F_3 = 5.38$
Innerh. d. Z.	127.19	96	1.32	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Differierten im 5. Schuljahr die Einstufungen wenig ängstlicher Schüler und sehr ängstlicher auf der Dimension S überzufällig, wobei die zuerst genannte Gruppe den informellen Test mit den Adjektiven 'hart', 'stark' und 'kraftvoll' einschätzte und die sehr ängstlichen Schüler ihn mit den Attributen 'weich', 'nachgiebig' und 'zart' charakterisierten, so fehlt diese Polarisierung der Urteile im 6. Schuljahr. Das Konzept 'informeller Test' wird in die Nähe des neutralen Mittelpunktes plaziert ($\bar{x}(W\bar{A}/KT)=3.7$ (3.7), $\bar{x}(W\bar{A}/KA)=3.9$ (3.6), $\bar{x}(S\bar{A}/KT)=3.9$ (4.1) und $\bar{x}(S\bar{A}/KA)=3.9$ (4.5)).

Tabelle 76: Tafel der Varianzanalyse: Konzept XVI/S

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	73.94	99		
Zw. d. Prüf-instrumenten	0.38	1		$F_1 = 0.51$
Zw. WÄ und SÄ	0.51	1		$F_2 = 0.68$
Ww.	0.47	1		$F_3 = 0.62$
Innerh. d. Z.	72.58	96	0.76	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Die Mittelwerte der Versuchsgruppen des 6. Schuljahrs tendieren auf der Dimension A, im Gegensatz zu denen des 5. Jahrgangs, zum negativen Bereich der Skala: $\bar{x}(W\bar{A}/KT)=4.0$ (4.1), $\bar{x}(W\bar{A}/KA)=3.6$ (4.0), $\bar{x}(S\bar{A}/KT)=3.9$ (4.0) und $\bar{x}(S\bar{A}/KA)=3.8$ (4.3). Kein F-Wert der Tabelle 77 ist signifikant.

Tabelle 77: Tafel der Varianzanalyse: Konzept XVI/A

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	103.26	99		
Zw. d. Prüf-instrumenten	1.82	1		$F_1 = 1.74$
Zw. WÄ und SÄ	0.05	1		$F_2 = 0.05$
Ww.	0.90	1		$F_3 = 0.86$
Innerh. d. Z.	100.49	96	1.05	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

5.4.9. Zusammenfassung der Ergebnisse

Vorangestellt sei, daß sich die Schüler des 6. Jahrgangs bei ihren Einstufungen differenzierter verhalten als ihre jüngeren Schulkameraden. Dieses Ergebnis konnte nach der allgemeinen Differenzierungshypothese erwartet werden. Einen Beleg dafür liefern erstens die im allgemeinen höheren F-Werte und zweitens die Wechselwirkungen, die bei den Konzepten 'Klassenarbeit', 'Zensur' und 'Klassentest' auftraten. Auch die Daten dieser Versuchsgruppe legen nahe, daß bestimmte Stereotype so fixiert sind, daß die relativ kurze Zeitdauer der experimentellen Bedingungen keine Versuchseffekte bewirkt. Dies müßte bei weiteren Untersuchungen berücksichtigt werden.

1. Ebenso wie bei den Schülern des 5. Schuljahrs ist das Selbstbild der älteren Schüler positiv gefärbt, doch stufen sich wenig ängstliche Schüler der Tendenz nach höher ein als sehr ängstliche ($F=2.17$ (0.07)). Das Selbstbild sehr ängstlicher Schüler wird auf der Dimension S durch die Art der Leistungsüberprüfung determiniert. Die Attribute 'hart', 'stark' und 'kraftvoll' werden nach dem informellen Test gewählt, während die Klassenarbeit weniger aggressive Selbsteinschätzungen auslöst, denn die Eigenschaften 'weich', 'nachgiebig' und 'zart' werden dem Konzept 'Selbstbild' zugeordnet. Einen auffällig hohen Grad an 'Eiligkeit', 'Aufgeregtheit' usw. bescheinigen sich wenig ängstliche Schüler nach dem informellen Test auf der Dimension A. Diese Tendenz spiegelt der relativ hohe F-Wert wider ($F=2.12$). Auch die Gruppe SA/KT des 5. Schuljahrs weist das gleiche Einstufungsverhalten auf.

2. Einen relativ hohen Effekt hat die Art des Prüfinstrumentes auf die Wahl der gewünschten Eigenschaften ($F=2.82ns$). Bedeutsame Unterschiede ($F=9.62ss$) bestehen zwischen wenig ängstlichen und sehr ängstlichen Schülern auf der Dimension B. Ein geringer Ängstlichkeitsgrad vergrößert den Wunsch nach positiven Eigenschaften. Dagegen sind sehr ängstliche Schüler zurückhaltender. Diese Relation konnte auch bei den Einstufungen des Konzepts Selbstbild beobachtet werden. Auf der Dimension S stufen sich wenig ängstliche und sehr ängstliche Schüler nach dem Klassentest als 'hart', 'stark' und 'kraftvoll' ein, nach der Bedingung Klassenarbeit jedoch als 'weich', 'nachgiebig' und

'zart'. Parallel dazu verlaufen die Einschätzungen auf der Dimension A ($F=8.45ss$). Wenn nun die Eigenschaften der Dimensionen S und A valide sind für das Bild des sich durchsetzenden, in Konkurrenz zu seinen Mitschülern stehenden Schülers, dann scheint der informelle Test wenig erwünschte Verhaltensweisen zu begünstigen, wenn man als pädagogische Zielvorstellung den sich mit seinen Klassenkameraden solidarisierenden, zur Zusammenarbeit fähigen Klassenkameraden vor Augen hat. Objektivisierte Leistungsmessung, die dumme und kluge Schüler produziert, hat dann dissoziative Funktionen.

3. Die Einstufungen auf den Dimensionen B, S und A werden weder durch den Grad der Ängstlichkeit noch durch die Art des Prüfinstruments affiziert. Das Bild des schwachen Schülers wird auf den Dimensionen B und A auf die neutrale Skalenmitte projiziert, was keineswegs einem stereotypen Verhalten entspricht. Auf der Dimension S jedoch wird mit den Adjektiven 'hart', 'stark' und 'kraftvoll' das Stereotyp 'schwacher Schüler' skizziert.

4. Die älteren Schüler scheinen beim Konzept 'guter Schüler' auf der Dimension B weniger stereotyphaft ihre Einstufungen vorzunehmen als die jüngeren. Dadurch wird auch der Effekt der Prüfinstrumente größer ($F_2=3.01$). Die Tendenz, daß der informelle Test im Durchschnitt niedrigere Einstufungswerte bewirkt, könnte erklärt werden mit den durch ihn ausgelösten Präferenzen für 'Härte', 'Stärke' usw., aggressiven Verhaltensweisen, die auch das Konzept 'guter Schüler' nicht unangetastet lassen. Auf der Dimension A ordnen sehr ängstliche Schüler dem Konzept ein Mehr an 'Lebhaftigkeit', 'Eiligkeit' usw. zu als die wenig ängstlichen. Doch auch dieser Unterschied ist nicht signifikant.

5. Die Urteile der Schüler für das Konzept 'Englischlehrer' fallen auf der Dimension B so eindeutig positiv aus, daß sie weder durch die Art des Prüfinstruments noch durch den Grad der Ängstlichkeit verändert werden. Dagegen rufen die Prüfinstrumente auf der Dimension S signifikante Effekt hervor. Das Konzept 'Englischlehrer' wird nach der Bedingung 'informeller Test' signifikant weniger positiv eingeschätzt als nach der Klassenarbeit. Die gleiche Tendenz hatte sich auch bei dem Konzept 'guter Schüler' gezeigt. Nach dem informellen Test halten

die Schüler den Englischlehrer für eiliger, lebhafter und reger als nach der Klassenarbeit ($F=2.17ns$).

6. Ein fast triviales Ergebnis liegt bei der Einstufung des Konzepts 'Klassenarbeit' auf der Dimension B durch sehr ängstliche Schüler vor, welche die Klassenarbeit schreiben mußten. Sie beurteilen das Konzept signifikant negativer als ihre weniger ängstlichen Mitschüler. Die Bedingung 'informeller Test' bleibt ohne Wirkung, obwohl man hätte erwarten können, daß die situativen Komponenten des Tests, der Zeitdruck und das Unterbinden von Rückfragen, eine positivere Bewertung der herkömmlichen Klassenarbeit hervorrufen. Im 5. Schuljahr hatte der Test negative Valenzen des Konzepts 'Klassenarbeit' evoziert. Es scheint, daß eine Leistungsüberprüfung durch hoch strukturierte Aufgabenformen für sehr ängstliche Schüler günstiger ist. Unser Untersuchungsergebnis kann der Hypothese SARASONS subsumiert werden: "Die Reaktion des prüfungsängstlichen Kindes auf Prüfungen und prüfungsähnliche Situationen in der Klasse spiegelt seine Erfahrungen in psychologisch oder interpersonal ähnliche Situationen in seinem Zuhause wider, vor und nach Beginn der formellen Einschulung."¹) Zwischen Prüfungsart und Erziehungsstil postulieren wir den Zusammenhang, daß rigide, autoritäre Erziehungsmaßnahmen mit der Struktur von Tests kongruieren. Ziehen wir zur Stützung unserer Hypothese die Vermutung zweier Autoren hinzu, die einen direktiven Erziehungsstil in Beziehung setzen zum Erleben von Angst: "Es wäre allerdings auch denkbar, daß Schüler von abweisenden, sich distanzierenden Lehrern tatsächlich weniger Angst erleben, da sie in einer fest strukturierten Umwelt leben. Denn das Verhalten dieser Lehrer dürfte dem ähnlich sein, das sie von anderen Erziehungspersonen (vor allem Eltern, aber auch anderen Lehrern) gewohnt sind. Bestehen dagegen erhebliche Diskrepanzen zwischen den Erziehungsstrategien der Eltern und denen des Lehrers, könnte das Kind eher Schwierigkeiten haben, sich in die Schulsituation einzufügen ... die sich dann in vermehrter Unsicherheit und Angst äußern."²) Die pädagogische Kon-

1) SARASON, S.B. u.a., a.a.O., S. 22

2) NICKEL, H./SCHLÜTER, P.: Angstwerte bei Hauptschülern und ihr Zusammenhang mit Leistungs- sowie Verhaltensmerkmalen, Lehrerurteil und Unterrichtstil, in: Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 1970, S. 134

sequenz besteht nun aber unserer Meinung nach nicht in der Propagierung informeller Tests, denn dies würde nur das Fortbestehen der kindlichen Konflikte unterstützen. Überdeckung der Symptome und nicht Beseitigung der Ursachen wäre die Folge.

7. Nach der Bedingung 'Klassenarbeit' beurteilen sehr ängstliche Schüler das Konzept 'Zensur' auf der Dimension B signifikant weniger positiv als ihre weniger ängstlichen Schulkameraden. Auch dieses Ergebnis paßt in die soeben gegebene Deutung. Die relativ unstrukturierte Situation 'Klassenarbeit' wird bei hohem Ängstlichkeitsgrad als etwas Bedrohliches empfunden, so daß die Einstufungen in die Richtung des negativen Skalenbereichs tendieren. 'Psycho'-logisch urteilen die sehr ängstlichen Schüler auch auf der Dimension S. Die Klassenarbeit mutet sie kaum noch 'hart', 'stark' und 'kraftvoll' an, sondern eher 'weich', 'nachgiebig' und 'zart' ($F=3.38_{ns}$).

8. Zwischen Prüfungsart und Ängstlichkeitsgrad besteht eine Wechselwirkung auf der Dimension B. Wenig ängstliche Schüler beurteilen nach der Klassenarbeit das Konzept 'Klassentest' positiver als sehr ängstliche. Zieht man zum Vergleich den Mittelwert der Einstufungen sehr ängstlicher Schüler nach der Bedingung 'informeller Test' heran, so scheint auch hier die Hypothese bestätigt, daß ein hoher Ängstlichkeitsgrad mit der Präferenz für hoch strukturierte Aufgabendarbietungen korreliert. Sehr ängstliche Schüler stufen nach dem Klassentest diese Prüfungsart höher ein als wenig ängstliche.

5.5. Die Prüfungsbedingungen bewirken keine Einstellungsänderungen bei wenig ängstlichen oder sehr ängstlichen Schülern

Wie die Tabellen 78 und 79 zeigen, bewirken die Prüfungsbedingungen weder im 5. noch im 6. Schuljahr Einstellungsänderungen. Damit wird die Hypothese (11) bestätigt: Der Einstellungsscore wird durch die Prüfung nicht beeinflusst.

Fast signifikant dagegen ist der Unterschied in den Einstellungen wenig ängstlicher und sehr ängstlicher Schüler des 6. Schuljahrs: Sehr ängstliche Schüler weisen im Durchschnitt eine positivere Einstellung gegenüber England, den Engländern und der englischen Sprache auf.

Tabelle 78: Tafel der Varianzanalyse: Einstellung gegenüber England, den Engländern und der englischen Sprache im 5. Schuljahr

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	16.66	99		
Zw. d. Prüf-instrumenten	0.02	1		
Zw. WÄ und SÄ	0.00	1		
Ww.	0.02	1		
Innerh. d. Z.	16.63	96	0.17	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

Tabelle 79: Tafel der Varinanzanalyse: Einstellung gegenüber England, den Engländern und der englischen Sprache im 6. Schuljahr

Variabilität	SAQ	FG	MQ	F-Wert
Gesamt	13.88	99		
Zw. d. Prüf-instrumenten	0.44	1		$F_1 = 3.14$
Zw. WÄ und SÄ	0.00	1		
Ww.	0.02	1		
Innerh. d. Z.	13.30	96	0.14	

$$F_{0.05;1/96}=3.94 \quad F_{0.01;1/96}=6.91$$

6. ZUSAMMENFASSUNG

Eine Prüfungsforschung, die sich hauptsächlich mit methodeninnernen Problemen wie z.B. der Verbesserung testtheoretischer Gütekriterien beschäftigt, hat zwar 'die Teile in der Hand, fehlt leider! nur das geistige Band! Sinn und Zweck des 1. Kapitels war es, den Zusammenhang der einzelnen Elemente nicht aus dem Auge zu verlieren. Die Thematisierung der gesellschaftlichen Rahmenbedingungen wies auf die Interdependenz von Prüfung, Ausbildung und Zensur in der Waren produzierenden Gesellschaft hin und hob im Anschluß an KROPE fünf Funktionen normorientierter Leistungsmessung hervor: Selektion, Arbeitsintensivierung, Konformitätsdruck, ideologische Integration und Überprüfung und Nachweis eines geforderten durchschnittlichen Qualifikationsniveaus.

Das 2. Kapitel zielte auf die wissenschaftstheoretische Problematisierung dreier Bereiche: auf das Validitätsproblem, die Relevanz positivistischer Forschungsdesigns vs. action research und auf den Stellenwert induktiv-statistischer Systematisierungen. Dabei wurde deutlich, daß das Gültigkeitsproblem nicht durch Verfahren gelöst werden kann, weil vor der Anwendung statistischer Verfahren die hermeneutische Identifikation der gültigen items geleistet werden muß, deren Zuordnung zu einem bestimmten Konstrukt durch ein gesellschaftlich bestimmtes Vorverständnis gesteuert wird. Als Entscheidungskriterium für die Auswahl eines Handlungs- oder eines empirisch-analytischen Ansatzes wurde die Problemstellung vorgeschlagen (vgl. Kap. 4, S.26). Die Verwendung induktiv-statistischer Systematisierungen wurde möglich durch eine Reinterpretation des Syllogismus, wobei nicht mehr die Konklusio als solche qualifiziert wurde, sondern die zwischen den Prämissen und der Konklusio bestehende Relation ausgezeichnet wurde. Probleme der Methodik wurden nur angedeutet, da sie unter 4.5. für die verwendeten Untersuchungsinstrumente expliziert werden sollten.

Die Problemstellung (Kapitel 3) richtete sich auf die Nebenwirkungen normorientierter Prüfinstrumente. Das Forschungsinteresse konzentrierte sich auf die emotionalen Komponenten von Einstellungen nach einer Prüfungssituation im Fach Englisch. Sehr ängstliche und weniger

ängstliche Schüler stufen die Konzepte 'Selbstbild', 'Wunschbild', 'schwacher Schüler', 'guter Schüler', 'Engischlehrer', 'Zensur' und 'Klassentest' ein. Die negativen Einstufungen auf den polar angelegten Adjektivskalen bezeichneten wir als dissoziativ, die positiven als integrativ. Die deskriptive Funktion der Versuchsergebnisse wurde betont; denn es handelt sich bei der vorliegenden Untersuchung um eine Pilotstudie.

Im 4. Kapitel wurden das Forschungsdesign, die Methoden und das statistische Instrumentarium dargestellt und die damit zusammenhängenden methodischen und methodologischen Probleme diskutiert. Die mit der Anwendung empirisch-analytischer Methoden speziell im pädagogischen Feld verbundenen Probleme wurden referiert. Der Einsatz des empirisch-analytischen Instrumentariums wurde von der Problemstellung her begründet. Nach grundsätzlichen Überlegungen zum Problem der adäquaten Methode für den zu untersuchenden Gegenstand wurden folgende Teilbereiche erörtert:

- 4.1. Zum Begriff des pädagogischen Experiments: Die Reichweite des Problems wurde abgesteckt durch die Begriffe 'technologisches' und 'praktisches' Experiment. Vom Horizont des technologischen Zugriffs hoben sich ab:
- das Postulat der Strukturidentität des Erklärungsschemas für naturwissenschaftliche und historische Wissenschaften mit dem Schlüsselbegriff 'nomologische Hypothese'.
 - die Konzentration auf methodenimmanente Problem wie z.B. Bildung von Kontrollgruppen, innere Validität usw.
 - die Reduktion intentionalen Handelns auf reaktives Verhalten, so daß Lernprodukte mehr interessieren als Lernprozesse.
 - die naive Gleichsetzung der inhaltlichen Bedeutungsanalyse mit dem praktischen Bedeutungskreis sprachlicher Ausdrücke, so daß z.B. die Interpretationsleistung einer Vp höchstens als Störvariable thematisiert wird.

Das praktische Experiment als Alternative transzendierte den technologischen Horizont durch

- den Hinweis auf die historische Bedingtheit sogenannter Quasi-invarianzen und die Kritik des Allquantors
- einen praktischen Erfahrungsbegriff als Alternative zur kontrollierten Erfahrung
- den Nachweis, daß das technologische Experiment nur den Widerspruch partikularer Interessen einebnet, sie aber nicht als veränderbare Variable begreift, während das praktische Experiment eine bessere Zukunft antizipiert, ihre Möglichkeit, und somit als regulative Idee die Wirklichkeit prüft.

- 4.2. Untersuchungshypothesen: Mit Hilfe der prädikatenlogischen Schreibweise wurden drei Arten von Hypothesen dargestellt. Sinn der Formalisierung war es nachzuweisen, daß der globale Vorwurf, empirische Psychologie oder Pädagogik abstrahiere von der Individualität des Schülers, nicht für alle Arten von Hypothesen zutrifft. Dem Charakter unserer Untersuchung als Pilotstudie wurde durch die Unterscheidung von 'offenen' und 'geschlossenen' Hypothesen analog zum Erkundungs- oder Entscheidungsexperiment Rechnung getragen. Die geschlossenen Hypothesen H (1) bis H (11) beziehen sich auf das Beurteilungsverhalten sehr ängstlicher und wenig ängstlicher Schüler gegenüber den Konzepten auf der Dimension (B) des Semantischen Differentials. Die Hypothesen H (1) bis H (3) wurden aufgrund der Annahme gebildet, daß die Wahrnehmungsdifferenzierung eine Funktion des Alters sei, somit ein sich signifikant unterscheidendes Urteilsverhalten zwischen den Schülern des 5. und 6. Schuljahrs zu erwarten sei. Der Einfluß der standardisierten, keine Rückfragen erlaubenden Instruktion der informellen Tests gegenüber der entspannteren Klassenarbeitsatmosphäre sollte mit den Hypothesen H (4) bis H (7) erfaßt werden. H (8) bis H (10) bezogen sich auf die Beurteilungseffekte, die durch hoch oder wenig strukturierte Prüfungssituationen momentan ausgelöst werden. Eine Kontrollfunktion sollte dagegen die Hypothese H (11) erfüllen, mit der die überdauernde Einstellung, das Einstellungstereotyp gegenüber den Engländern, England und der englischen Sprache gemessen werden sollte. Die unterschiedlichen Prüfungsbedingungen, so wurde vermutet, würden keinen Einfluß auf den Einstellungsscore haben. Die 'offene' Hypothese H (12) sollte die emotionalen Reaktionen gegenüber den acht Konzepten auf den Dimensionen (S) und (A) des Semantischen Differentials erfassen.
- 4.3. Der Versuchsplan: Das Forschungsdesign sollte optimal dem Untersuchungsgegenstand angepaßt werden: der restringierten Prüfungssituation, in der das Subjekt als Objekt behandelt wird. Damit wurde aufgrund der Problemstellung gegen die Alternative Handlungsforschung entschieden. Den Versuchsbedingungen entsprach der unvollständige, zweifaktorielle Versuchsplan.

- 4.4. Die experimentellen Gruppen sind nicht repräsentativ für eine Grundgesamtheit. Das Kriterium ist jedoch nicht unbedingt erforderlich, wenn Aussagen über Wirkungszusammenhänge zwischen abhängigen und unabhängigen Variablen (s.S. 48) gemacht werden sollen. Den 200 Schülern des 5. und 6. Schuljahrs wurden ein Angst-, Fremdspracheneignungs- und Intelligenztest administriert. Die Variable Alter wurde ebenfalls kontrolliert. Die Versuchsgruppen unterschieden sich nicht signifikant in den Ausgangsbedingungen.
- 4.5. Für die Überprüfung der Hypothese H (11) mußte ein Meßinstrument entwickelt werden, die Einstellungsskala 'Englisch'. Die grundsätzliche methodologische Kritik an der Einstellungs- und Meinungsforschung, der Vorwurf, situations-invariante Einstellungsmessung zu präferieren, schlug für unsere Zwecke ins Positive um, da wir stereotype Einstellungen erfassen wollten. Mit Hilfe des Paarvergleichs wurde einmal eine indikatororientierte (50 Studenten) und zum anderen eine reaktionsorientierte (153 Schüler) Skalierung der ursprünglich 14 Behauptungen durchgeführt. Die Reliabilitätskoeffizienten streuen von $r = 0.69$ bis $r = 0.91$. Der Validitätskoeffizient - die Rangkorrelation von Schüler- und Studentenurteil - beträgt nach Elimination der items 4 und 14 $r_g = 0.948$.
- Adjektivpaare und Skalen des Semantischen Differentials übernahmen wir für unsere Konzepte von HURRELMANN, der faktorenanalytisch nachgewiesen hat, daß die Skalen auf den Dimensionen Bewertung, Stärke und Aktivität laden. Das hat den Nachteil, daß die Adjektivpaare für die Konzepte nicht aufgrund einer phänomenologischen Bedeutungsanalyse ausgewählt worden sind.
- Mit der Übernahme der tiefenpsychologisch fundierten habit-Interferenz-Theorie von MANDLER und SARASON fiel auch die Entscheidung für die von den genannten Autoren entwickelten Angstskalen, jedoch in der von NICKEL u.a. für deutsche Verhältnisse adaptierten Form. Die Fremdspracheneignung wurde mit dem FTU 4 - 6, die Intelligenz mit dem LPS erfaßt. Dabei wurden die Grundannahmen der klassischen Testtheorie problematisiert. Bei den von uns konstruierten informellen Tests I und II wurden sprachwissenschaftliche, psychologische und didaktische Kriterien berücksichtigt.

5. Untersuchungsergebnisse und Ausblick: Wenn wir auch die experimentellen Gruppen nicht als repräsentative Stichproben konstruiert hatten, so nahmen wir dennoch an, daß die Varianz der Merkmalsausprägungen der abhängigen Variablen auf die experimentellen Bedingungen zurückzuführen seien. Exemplarisch konnte dies an der generellen Hypothese (S.101) nachgewiesen werden: "Die Übereinstimmung (Korrelation) der Statistiken, die zu Lasten der experimentellen Bedingung 'informeller Test' geht, muß größer sein als die Übereinstimmung der Statistiken, die zu Lasten der experimentellen Bedingung 'Klassenarbeit' geht." Der vermutete Zusammenhang gründete auf der Annahme der größeren Prägnanz der Teststruktur gegenüber der herkömmlichen Klassenarbeit; sie wurde abgeleitet aus dem Figur-Grund-Prinzip der Wahrnehmungspsychologie. Wie erwartet, traf die größere Korrelation für die Testeinstufungen ($r_{KT-5/KT-6}=0.88$), die niedrigere für die Klassenarbeit ($r_{KA-5/KT-6}=0.61$) ein. Damit scheint die Generalhypothese, daß der Strukturierungsgrad normorientierter Prüfinstrumente unterschiedliche Einstufungseffekte erzeugt, bestätigt.

H (1): "Schüler des 5. Schuljahres unterscheiden sich signifikant in der Beurteilung der Konzepte von den Schülern des 6. Schuljahres."

Zur Überprüfung der 'geschlossenen' Hypothese wurde die Prüfgröße t nach dem t -Test für unabhängige Stichproben auf zwei Arten berechnet:

1. wurden die Einstufungen der Konzepte nach unterschiedlichen Prüfungsbedingungen auf Signifikanz geprüft. Überzufällige Unterschiede zeigten sich bei den 'personbezogenen' Konzepten X (Wunschbild), XI (schwacher Schüler), XII (guter Schüler) und XIII (Englischlehrer):

$$t_{KT-5/KT-6;X} = 2.56s \quad df = 98$$

$$t_{KT-5/KT-6;XI} = 3.25s \quad df = 98$$

$$t_{KT-5/KT-6;XII} = 3.35s \quad df = 98$$

$$t_{KT-5/KT-6;XIII} = 2.31s \quad df = 98$$

$$t_{KA-5/KT-6;XIII} = 2.73s \quad df = 98$$

2. wurden die Prüfgrößen unter Vernachlässigung der Prüfungsbedingungen für $df=198$ Freiheitsgrade berechnet:

$$t_{5/6;IX-B}=2.33s, \quad t_{5/6;X-B}=0.88, \quad t_{5/6;XI-B}=1.85, \quad t_{5/6;XII-B}=0.87, \quad t_{5/6;XIII-B}=2.62s, \quad t_{5/6;XIV-B}=1.53, \quad t_{5/6;XV-B}=0.69$$

und $t_{5/6;XVI-B}=0.55.$

H (2): "Schüler des 5. Schuljahres unterscheiden nicht zwischen den Konzepten KLASSENARBEIT und KLASSENTEST."

Die Hypothese wurde bestätigt. Der t-Wert für korrelierte Stichproben ergab die Prüfgröße $t=1.54ns$ bei $df=99$ Freiheitsgraden.

H (3): "Schüler des 6. Schuljahres unterscheiden zwischen den Konzepten KLASSENARBEIT und KLASSENTEST."

Die Schüler diskriminieren sehr signifikant zwischen beiden Konzepten: $t_{\text{korr.}}=4.02ss$. Sowohl die Bedingungen 'Prüfinstrument' als auch der Ängstlichkeitsgrad erzeugten signifikante Effekte: $t_{XIV/XVI;KT}=3.13ss$, $t_{XIV/XVI;KA}=2.46s$, $t_{XIV/XVI;WA}=2.75ss$ und $t_{XIV/XVI;SA}=2.92ss$.

H (4): "Aufgrund der experimentellen Bedingung 'Klassentest' wird das Konzept 'ENGLISCHLEHRER' von sehr ängstlichen Schülern negativ bewertet."

Weder die differentielle Analyse der Daten des 5. noch des 6. Schuljahres erbrachte eine Bestätigung der Hypothese.

H (5): "Ein Prüfinstrument, das weniger 'autoritär' anmutet, wird von sehr ängstlichen Schülern positiv beurteilt".

Die sehr ängstlichen Schüler des 5. Schuljahres, die der Bedingung Klassenarbeit unterworfen waren, beurteilten das Konzept Klassenarbeit positiver als das Konzept Klassentest, so daß die Hypothese für diesen Jahrgang zutraf: $t_{XIV/XVI;SA}=8.97ss$.

H (6): "Die sehr ängstlichen Schüler des 6. Schuljahres werden das Konzept KLASSENTEST negativer beurteilen als die Schüler des 5. Schuljahres."

Die Hypothese konnte nicht bestätigt werden: $t = 0.97ns$.

H (7a): "Aufgrund der experimentellen Beeinflussung 'Klassentest' wird das Konzept ZENSUR von sehr ängstlichen Schülern negativ beurteilt."

Auch diese Vermutung traf nicht ein.

H (7b): "Die sehr ängstlichen Schüler des 6. Schuljahres werden das Konzept ZENSUR negativer beurteilen als die Schüler des 5. Schuljahres."

Der t-Wert war nicht signifikant.

H (8): "Sehr ängstliche und wenig ängstliche Schüler unterscheiden sich in ihrem Selbstbild im Anschluß an verschiedene Prüfungssituationen."

Weder im 5. noch im 6. Schuljahr konnte die Hypothese bestätigt werden.

H (9): "Wunschkild und Selbstbild stimmen nach hoch strukturierten Prüfungssituationen mehr überein als nach wenig strukturierten."

Wie aus der Tabelle 80 zu ersehen ist, konnte die Vermutung nicht bestätigt werden.

Tabelle 80: Mittelwerte der Konzepte IX und X auf der Dimension B für das 5. und 6. Schuljahr

		WÄ/KT	SÄ/KT	WÄ/KA	SÄ/KA
5. Sch.:	IX	5.3	5.1	5.2	5.3
	X	6.1	6.4	5.9	5.7
6. Sch.:	IX	5.4	5.2	5.6	5.3
	X	6.3	5.8	6.1	5.5

Hoch strukturierte Prüfungssituationen scheinen eher die Diskrepanz zwischen Selbst- und Wunschkild zu vergrößern. Diese Vermutung müßte jedoch in einer weiteren Untersuchung überprüft werden.

H (10): "H (9) tritt schärfer bei Schülern des 6. Schuljahres hervor."

Auch diese Hypothese konnte nicht generell bestätigt werden.

Die von uns vorgelegte Untersuchung, welche affektive Wirkungen normorientierter Prüfinstrumente als verbales Verhalten erfaßte, wie es sich in den Einstufungen ausgewählter Konzepte niederschlug, müßte ergänzt werden durch Forschungsarbeiten, die overttes Verhalten einbeziehen. Kontrastiv zum normorientierten Ansatz müßten die Effekte des lernzielorientierten Ansatzes zum Gegenstand weiterer Untersuchungen gemacht werden. Beide Ansätze müßten auch (die hier nur angedeuteten) innerschulischen Rahmenbedingungen in ihre Fragestellungen aufnehmen:

- die Transparenz vs. Nichttransparenz von Lernzielen
- den Grad der Steuerung von Lernprozessen
- verschiedene Lernzielebenen
- einen angstfreien vs. angstinduzierenden Lehrstil
- die Kommunikationsstruktur (reversibel/irreversibel)
- einen stofforientierten vs. schülerorientierten Unterricht
- das Konfliktverhalten von Schülern im Anschluß an sich unterscheidende Prüfungssituationen.

ANHANG

1. INFORMELLER TEST I` (5tes Schuljahr)

1.1. Anweisungen für die Prüfung (Tonbandtext)

Informeller Test I

Nachdem Ihr nun alle Euren Namen, den Vornamen und das Datum eingetragen habt, können wir beginnen.
Blättert bitte das Titelblatt um!

SUBTEST I

Wir beginnen mit der Übung "Lautdiskrimination". Zu dieser Übung gehören 16 Aufgaben. Nach jeder Aufgabe stehen die Zahlen eins, zwei und drei. Zu jeder Zahl gehört ein Wort. Von den jeweils drei Wörtern, die zu einer Aufgabe gehören, klingen zwei gleich. Nur eins klingt anders. Dieses anders klingende Wort müßt Ihr anstreichen. Es kann an erster, zweiter oder dritter Stelle stehen.

Ich gebe Euch ein Beispiel: house - mouse - house.

Das anders klingende Wort steht an zweiter Stelle. Also hättet Ihr die Zwei anstreichen müssen.

Hört noch ein Beispiel: can - can - pan.

Hier hättet Ihr die Drei anstreichen müssen, denn das anders klingende Wort "pan" steht an dritter Stelle.

Jetzt beginnen wir mit den sechzehn Aufgaben, die auf der oberen Blatthälfte stehen.

- | | |
|-------------------------------|--------------------------------|
| Aufgabe (1) that - that - fat | (9) day - play - play |
| (2) hat - that - hat | (10) my - eye - my |
| (3) is - his - his | (11) long - song - song |
| (4) thin - in - in | (12) do - do - two |
| (5) close-nose - close | (13) Ted - Ted - red |
| (6) now - how - now | (14) mister - sister - sister |
| (7) dark-park - dark | (15) mother - mother - brother |
| (8) fair - hair - fair | (16) not - hot - hot |

SUBTEST II

Wir kommen nun zur Übung "Hörverständnis I" links unten auf Eurem Blatt. Dazu gehören 10 Aufgaben. Jede Aufgabe besteht aus einem Satz, der mit "yes" oder "no" beantwortet werden kann. Zum linken Kästchen gehört jeweils das "yes", zum rechten Kästchen das "no". Ihr werdet jeden Satz nur einmal hören, und Ihr müßt entscheiden, ob der Satz einen richtigen Sinn ergibt. Dazu einen Beispielsatz:

"Your teacher is biting the dog."

Das kommt nur in den seltensten Fällen vor, daß ein Lehrer einen Hund beißt. Ihr hättet also das "no"-Kästchen ankreuzen müssen.

Also aufgepaßt! Aufgabe:

- (1) The pond is in the house
- (2) You have a dog on your head
- (3) Your friend has three eyes
- (4) Mr. Barton is Tom's mother
- (5) Your teacher has an English book
- (6) Tom and Susan have a play-room
- (7) The Bartons have a car
- (8) You have eleven brothers
- (9) A shoe is in your pocket
- (10) The Bartons have a boat

SUBTEST III

Wir gehen gleich weiter zur nächsten Aufgabe. Rechts unten auf der Seite steht die Übung Hörverständnis (2), zu der auch zehn Aufgaben gehören. Neben jeder Aufgabe stehen drei verschiedene Zahlen. Ich nenne Euch gleich jeweils eine englische Zahl, die Ihr ankreuzen sollt.

Aufgabe: (1) four (2) eleven (3) ten (4) three (5) two
(6) one (7) five (8) seven (9) nine (10) six

SUBTEST IV

Wir blättern um zur nächsten Seite. Dort seht Ihr sechzehn Bilder. Seht Euch diese Bilder, ohne dabei zu sprechen, genau an! Ihr werdet nun gleich eine Anzahl englischer Wörter hören. Vor jedem Wort wird eine Zahl genannt - zum Beispiel (1) sun oder (8) ear. Wäre also in einem großen Kästchen irgendwo eine Sonne, dann müßtet Ihr in das zum Bild gehörende kleine Eckkästchen unten rechts eine 1 eintragen. Ein anderes Beispiel: (8) ear - dann müßtet Ihr eine 8 in das Kästchen des passenden Bildes eintragen. Bitte aufpassen! Wir beginnen:

Neun - eye, sechs - fish, sieben - car, elf - nose, zwei - pullover, elf - swimmer, acht - shoe, sieben - goose, acht - dog, eins - boy, fünf - water, drei - pocket, vier - door, drei - house, sieben - board, zehn - window.

SUBTEST V

Blättert um zur nächsten Seite! Ihr habt jetzt vier Aufgaben zu lösen. Ihr seht eine Mütze, die einmal auf dem Kasten, in ihm, hinter ihm oder unter ihm liegt. Neben jeder Zeichnung ist ein Satz. Ihr braucht immer nur das fehlende englische Wort einzutragen. Dafür habt Ihr zwei Minuten Zeit. (Nach zwei Minuten)
Halt! Bitte, die Bleistifte hinlegen und umblättern!

SUBTEST VI

Jetzt habt Ihr sechs Aufgaben zu lösen. Bei diesem Test - oben steht Wortschatz (3) - sollt Ihr in die Lücke eines jeden Satzes das passende Eigenschaftswort eintragen. Bitte beginnt!
Halt, die Bleistifte hinlegen und umblättern!

SUBTEST VII

Oben steht jetzt Wortschatz (4). Auch bei dieser Übung braucht Ihr nur ein Wort einzutragen. Das Wort gibt jeweils an, ob zum Beispiel ein Haus mein Haus ist, Euer Haus oder sein Haus ist. Tragt das passende Wort ein!
Halt! Geht jetzt weiter zur Übung Syntax (2) !

SUBTEST VIII

Diese Übung besteht aus zehn Aufgaben. Ihr habt jeweils die Fragen schriftlich zu beantworten. Also beginnt! Ihr habt dafür fünf Minuten Zeit.
Halt, Bleistifte hinlegen und umblättern!

SUBTEST IX

Wir kommen jetzt zur letzten Seite. Euer Lehrer wird Euch die Aufgabe erklären. Wir üben diese zehn Aufgaben an einem Beispiel. Euer Lehrer schreibt jetzt an die Tafel "Tom's is this hat". Unter "Tom's" steht eine 1, unter "is" eine 2, unter "this" eine 4, unter "hat" eine 5. Neben diesem verdrehten Satz zeichnet Euer Lehrer einen waagerechten Balken, den er in sieben Kästchen aufteilt, genauso wie es in Eurem Testheftchen ist. Jetzt müßt Ihr überlegen, wie der Satz richtig heißt. Die richtige Lösung ist: "This is Tom's hat".

In der gleichen Reihenfolge wie die Wörter im Satz stehen, müßt Ihr jetzt rechts die Zahlen eintragen. Der richtige Satz heißt : "This is Tom's hat". Unter "this" steht eine drei. Also müßt Ihr in das erste Kästchen rechts eine Drei eintragen. Unter dem "is" steht eine Zwei. Jetzt müßt Ihr die Zwei eintragen. Der Satz hieß: "This is Tom's hat". Unter "Tom's" steht eine eins, also müßt Ihr jetzt eine Eins eintragen, unter "hat" steht eine vier, also kommt jetzt eine Vier. Jetzt könnt Ihr an die Arbeit gehen. Für die übrigen Beispiele habt Ihr sechs Minuten Zeit.

(Nach 6 Minuten)

Halt!

SUBTEST X

Wir kommen jetzt zu den letzten acht Aufgaben. Hinter jedem Fragezeichen steht ein Antwortsatz. Eure Aufgabe ist es nun, zu dem Antwortsatz die dazugehörige Frage zu stellen und aufzuschreiben. Bitte, beginnt!

Halt, die Bleistifte hinlegen und die Testhefte zuklappen!
Recht herzlichen Dank!

1.2. Testbogen "INFORMELLER TEST I (5)"

NAME: _____ VORNAME: _____

Punkte	Dis- krimi- nation	Hör- ver- ständnis		Lexeme				Syntax				
	1	2	3	4	5	6	7	8	9	10	x	f.x
16												
15												
14												
13												
12												
11												
10												
9												
8												
7												
6												
5												
4												
3												
2												
1												
0												

Punkte: _____ Note: _____

LAUTDISKRIMINATION

(1)	1	2	3
(2)	1	2	3
(3)	1	2	3
(4)	1	2	3
(5)	1	2	3
(6)	1	2	3
(7)	1	2	3
(8)	1	2	3
(9)	1	2	3
(10)	1	2	3
(11)	1	2	3
(12)	1	2	3
(13)	1	2	3
(14)	1	2	3
(15)	1	2	3
(16)	1	2	3

BORVERSTANDNIS (1)

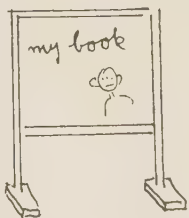
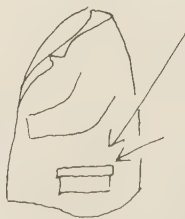
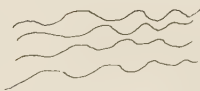
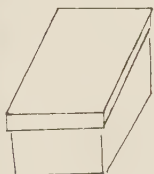
Yes No

(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)
(10)

BORVERSTANDNIS (2)

(1)	7	4	11
(2)	11	3	1
(3)	10	2	9
(4)	7	8	3
(5)	6	2	5
(6)	11	1	3
(7)	5	4	9
(8)	7	2	8
(9)	6	10	9
(10)	10	11	6

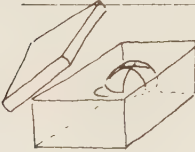
WORTSCHATZ (1)



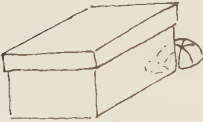
WORTSCHATZ (2)



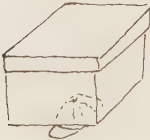
The cap isthe box



The cap isthe box



The cap isthe box



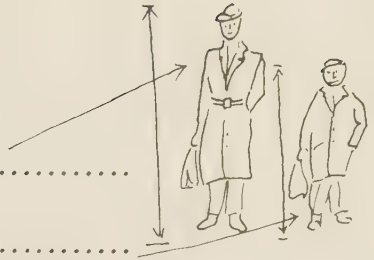
The cap isthe box

WORTSCHATZ (3)

(1)

This man is

This man is



(2)



This girl has hair

(3)



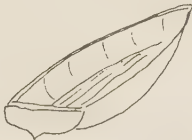
This boat is

(4)



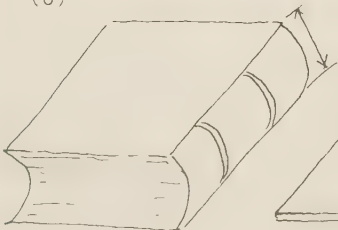
This boat is

(5)

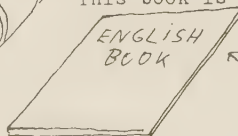


This boat is

(6)



This book is



This book is

WORTSCHATZ (4)

- (1) Tom has a cap. It iscap.
(2) Tom and Susan have a dog. It isdog.
(3) Susan has a bag. It isbag.
(4) I have a pen. It ispen.
(5) Father has a car. It iscar.
(6) You have a friend. It isfriend.
(7) Mother has a hat. It ishat.
(8) The girl has a dog. It isdog.
(9) Father has a new cap. It iscap.
(10) Father and mother have a new car. It iscar.

SYNTAX (1)

- (1)  Is it a house? _____
(2)  Is it a shoe? _____
(3) _____ Have you a pen? _____
(4) _____ Is Susan a girl? _____
(5)  Is this a box? _____
(6) _____ Has Susan a brother? _____
(7) _____ Has Tom a sister? _____
(8) _____ Is Mr. Barton Tom's father? _____
(9)  Is this a window? _____
(10) _____ Has Susan a brother? _____

SYNTAX (2)

- (1) show your me new bike!
 1 2 3 4 5
- (2) play let's a game!
 1 2 3 4 5
- (3) in his Tom a book has hand.
 1 2 3 4 5 6 7
- (4) a swimmer good Tom is.
 1 2 3 4 5
- (5) on his head has a cap Tom?
 1 2 3 4 5 6 7
- (6) she in her has pocket what?
 1 2 3 4 5 6
- (7) the is car where?
 1 2 3 4
- (8) Tom washing car is father's
 1 2 3 4 5
- (9) looking at the are board you?
 1 2 3 4 5 6
- (10) at head look my!
 1 2 3 4

SYNTAX (3)

- (1) _____ ?Tom is in
the water.
- (2) _____ ? Tom has a
cap on his head.
- (3) _____ ? The doors
are white.
- (4) _____ ? I have three books.
- (5) _____ ? Mr. Barton is
washing the car.
- (6) _____ ? The boat is blue.
- (7) _____ ? Susan is on
the bank.
- (8) _____ ? Yes, I have
a book!

2. INFORMELLER TEST II (6tes Schuljahr)

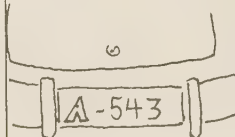
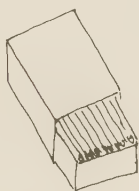
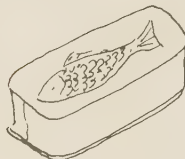
NAME: _____ VORNAME: _____ DATUM: _____

Punkte	Wortschatz				Grammatik			x	f.x
	1	2	3	4	1	2	3		
16									
15									
14									
13									
12									
11									
10									
9									
8									
7									
6									
5									
4									
3									
2									
1									
0									

Punkte: _____ Note: _____

WORTSCHATZ I

Unter jedem Bild ist eine freie Zeile. Trage dort das passende englische Wort ein!



WORTSCHATZ II

Fülle die Lücken mit dem passenden Wort aus!

1. In England they speak
2. In France they speak
3. In Italy they speak
4. In Sweden they speak
5. In Spain they speak
6. In Switzerland they speak
7. In Germany they speak
8. In Denmark they speak
9. In Holland they speak

WORTSCHATZ III

Setze in die Lücken something, anything oder nothing ein!

Is therein the bottle?

No, there isn't in it.

There isin it.

Is there in that glass?

Yes, there isin it.



WORTSCHATZ IV

Suche zu jedem Wort das Gegenteil!

Beispiel: thick - (thin)

- to forget -
- to find -
- to go away -
- to stop -
- to walk slowly -
- to reach -
- to take away -
- to ask -

GRAMMATIK I

Setze him, her oder them ein!

1. Tom need's a new coat. His mother buysa nice blue coat.
2. Susan can't open the window. Tom helps
3. Mary's friends give Mary nice presents. She thanksall.
4. Tom doesn't find the right shoes. The shop assistent shows
.....more shoes.
5. Poor Jane! Nobody can tellwhere Mrs Parker's home is.

GRAMMATIK II

Bilde zu den vorgegebenen Antworten die richtigen Fragen und schreibe sie auf die leer gelassene Linie!

1. _____ ? Mr. Pound
gets up at seven o'clock.
2. _____ ? In the afternoon
Mr. Pound reads and dictates letters.
3. _____ ? Every day the
managers dictate letters.
4. _____ ? Mrs. Brown
makes two dresses in a week.
5. _____ ? I write letters
at the office.
6. _____ ? Mrs. Brown
makes evening dresses.
7. _____ ? I read
English and German newspapers.
8. _____ ? Mr. Pound
spends his weekend in the country.

GRAMMATIK III

Setze die richtige Form des Tätigkeitswirtes ein!

1. At a quarter to five Tom and Susan (to come)at home.
2. First Susan (to carry)her schoolbag into her room.
3. Then she (to run)..... into the living-room.
4. Then she (to hurry) to her room.
5. Susan (to thank)her mother.
6. Then she (to put on) her coat and (to run)
away to the bus stop.
7. At the bus stop many people (to wait)for the bus.
8. Susan (to wait), too.

3. EINSTELLUNGSSKALA "ENGLISCH"

3.1. Instruktion für Schüler (Reaktionsskalierung)

In diesem Jahr beginnt für Euch der Englischunterricht. Mir geht es darum zu erfahren, was Ihr von England, den Engländern und der englischen Sprache haltet. Deshalb findet Ihr auf den folgenden Seiten eine Anzahl von Behauptungen zu dem genannten Bereich. Jeweils zwei Behauptungen sollt Ihr miteinander vergleichen. Am Rande kreuzt Ihr dann die Behauptung an, der Ihr am meisten zustimmt. Stimmenthaltungen sind nicht zugelassen.

BEISPIEL

- A Engländer sind sehr traditionsbewußt
- B Engländer sind im allgemeinen sehr fleißig

3.2. Instruktion für Studenten (Stimuluskalierung)

Da im deutschsprachigen Raum keine nach psychometrischen Methoden konstruierte Skala existiert, welche die Einstellung von Hauptschülern gegenüber England, den Engländern, der englischen Sprache, dem Englischunterricht zu erfassen erlaubt, diese aber meiner Meinung nach für den Englischlehrer wichtig ist, versuche ich, dies nachzuholen. Sie erhalten gleich von mir ein Testheft und einen Antwortbogen. Auf den 14 Seiten des Testheftes finden Sie jeweils Aussagenspaare. Sie fungieren als Experten und sollen - unabhängig von Ihrer eigenen Einstellung - entscheiden, welche Aussage des jeweiligen Aussagenspaares für die Engländer positiver ist. Urteile der Form A=B sind nicht zugelassen.

BEISPIEL

- A Engländer sind sehr traditionsbewußt
- B Engländer sind im allgemeinen sehr fleißig

3.3. Muster des 14-seitigen Fragebogens

- A Ich habe nichts gegen Engländer, doch glaube ich, daß sie ein wenig komisch sind.
- B Meine Ferien würde ich schon in England verbringen, doch möchte ich nicht für immer dort wohnen.
- A Ich habe nichts gegen Engländer, doch glaube ich, daß sie ein wenig komisch sind.
- C Mir genügte es, einen englischen Brieffreund (eine englische Brieffreundin) zu haben.
- A Ich habe nichts gegen Engländer, doch glaube ich, daß sie ein wenig komisch sind.
- D Ich kann mir gut vorstellen, mit einem Engländer (einer Engländerin) verheiratet zu sein.

3.4. Endform der Einstellungsskala "ENGLISCH" mit den Skalenwerten reaktions- und stimulusorientierter Messung

NAME: VORNAME:

Bitte lies Dir zuerst die 12 unten stehenden Behauptungen durch!
Dann kreuze die Behauptungen an, denen Du am meisten zustimmst.
Zum Schluß kennzeichne mit einem zusätzlichen Kreuz die Behauptung, der Du am meisten zustimmst.

Herzlichen Dank!

	Skalenwerte	
	rM	sM
1. Ich möchte die englische Sprache am liebsten in England lernen	1.380	2.683
2. Ich habe nichts gegen Engländer, doch glaube ich, daß sie ein wenig komisch sind	0.680	1.535
3. Ich kann Engländer nicht leiden	0.000	0.000
4. So sehr interessieren mich die Engländer, England und die englische Sprache auch wieder nicht, daß ich mich mit ihnen mehrere Jahre in der Schule beschäftigen möchte	0.770	1.287
5. Das Erlernen der englischen Sprache weitet den Blick des Menschen	1.630	2.490
6. Mir genügt es, einen englischen Brieffreund (eine englische Brieffreundin) zu haben	0.930	1.796
7. Ich kümmere mich überhaupt nicht um das, was mit England usw. zu tun hat	0.670	0.901
8. Englisch ist eine Weltsprache, so daß es sich lohnt, sie zu erlernen	1.600	2.638
9. Ich kann mir gut vorstellen, mit einer Engländerin (einem Engländer) verheiratet zu sein	0.800	2.483
10. Anstelle von Englisch sollte man heutzutage mehr Mathematik unterrichten	0.650	0.974
11. Meine Ferien würde ich schon in England verbringen, doch möchte ich nicht für immer dort wohnen	1.010	2.171
12. Meine Einstellung gegenüber England, den Engländern und der englischen Sprache ist eher gleichgültig	0.740	1.252

4. ANGSTSKALEN

4.1. Allgemeine Angstskala CMAS

1. Es fällt mir schwer, im Unterricht immer ganz bei der Sache zu sein.
2. Ich werde unruhig, wenn mich jemand bei der Arbeit beobachtet.
3. Ich bin meist unglücklich, wenn ich nicht alles besonders gut mache.
4. Manchmal habe ich starkes Herzklopfen.
5. Zuweilen ist mir zumute, daß ich weinen möchte.
6. Manchmal wünschte ich, daß ich weit von hier fort wäre.
7. Anderen scheint alles leichter zu fallen als mir.
8. Insgeheim habe ich vor einigen Dingen etwas Angst.
9. Ich merke zuweilen, daß andere die Art, wie ich mich verhalte, nicht mögen.
10. Manchmal fühle ich mich einsam, auch wenn ich mit anderen zusammen bin.
11. Ich kann mich nur schwer entscheiden, ob ich etwas tun oder lassen soll.
12. Ich werde leicht unruhig, wenn etwas nicht so geht, wie ich es mir eigentlich vorgestellt habe.
13. Ich mache mir zuviel Sorgen.
14. Ich mache mir oft Gedanken darüber, was meine Eltern wohl zu mir sagen werden.
15. Ich kann oft nicht ruhig atmen.
16. Ich ärgere mich oft über Kleinigkeiten.
17. Es gibt zuweilen Situationen, in denen merke ich, daß meine Hände anfangen, feucht zu werden.
18. Zuweilen muß ich häufig austreten gehen.
19. Ich habe das Gefühl, daß andere Kinder nicht so unglücklich sind wie ich.
20. Ich mache mir oft Gedanken darüber, was andere von mir halten.
21. Manchmal fällt mir das Schlucken schwer.
22. Oft habe ich vor etwas Angst, das sich hinterher als gar nicht so schlimm erweist.
23. Man kann meine Gefühle leicht verletzen.
24. Ich habe sehr oft Angst, daß ich nicht das Richtige tue.
25. Ich habe ein wenig Angst vor dem, was alles noch kommen kann.
26. Oft kann ich abends lange nicht einschlafen, weil ich mir zuviele Gedanken mache.
27. Ich mache mir oft Sorgen, ob ich in der Schule wohl auch alles richtig machen werde.

28. Ich empfinde es oft als ungerecht, daß gerade ich ausgeschimpft werde.
29. Oft fühle ich mich ganz allein. wenn ich mit Kameraden (innen) zusammen bin.
30. Manchmal merke ich den anderen an, daß jemand mir sagen will, daß ich etwas falsch mache.
31. Ich fürchte mich ein wenig, wenn es dunkel ist.
32. Es fällt mir manchmal schwer, im Unterricht mitzukommen.
33. Ich habe oft ein komisches Gefühl im Magen.
34. Wenn ich abends zu Bett gehe, mache ich mir oft Gedanken darüber, was am nächsten Tag alles kommen wird.
35. Es kommt vor, daß ich Dinge tue, von denen ich hinterher wünschte, ich hätte sie nie getan.
36. Ich habe oft Kopfschmerzen.
37. Zuweilen mache ich mir Sorgen darüber, daß meinen Eltern etwas zustoßen könnte.
38. Ich fühle mich oft so müde.
39. Oft habe ich Träume, die mich sehr beunruhigen.
40. Ich werde leicht nervös.
41. Oft mache ich mir Sorgen darüber, daß mir etwas Schlimmes zustoßen könnte.

4.2. Prüfungsangstskala TASC

42. Ich habe immer etwas Angst, wenn die Aufgaben abgefragt werden oder im Mündlichen geprüft wird.
43. Ich mache mir oft Sorgen, ob ich versetzt werde.
44. Wenn der Lehrer mich vor die Klasse ruft und ich vorlesen muß, habe ich immer Angst, daß ich mich verspreche.
45. Wenn der Lehrer jemanden vorn zur Tafel ruft, denke ich meistens: Hoffentlich nimmt er mich nicht.
46. Manchmal träume ich nachts, daß ich in der Schule bin und mich vor der Klasse blamiere, weil ich etwas nicht kann.
47. Wenn der Lehrer prüfen will, ob wir zu Hause gelernt haben, bekomme ich immer Herzklopfen.
48. Ich habe das Gefühl, daß andere Kinder meiner Klasse beim Englischunterricht besser mitkommen als ich.
49. Wenn ich abends im Bett liege, muß ich oft daran denken, ob ich am nächsten Tag in der Schule auch alles richtig machen werde.
50. Wenn ich vor der Klasse etwas an die Tafel schreiben muß, zittert meine Hand oft ein wenig.
51. Mir ist oft so, als würden meine Mitschüler im Unterricht leichter mitkommen als ich.

52. Ich habe das Gefühl, daß ich mir mehr Sorgen über die Schule mache als andere Kinder meiner Klasse.
53. Wenn ich zu Hause an den Englischunterricht vom nächsten Tag denke, bekomme ich manchmal Angst, daß ich eine falsche Antwort geben könnte.
54. Wenn ich krank bin und die Schule versäumen muß, quält mich der Gedanke daran, was ich alles versäume und daß meine Schulleistungen schlechter sein werden als die der anderen Kinder, wenn ich wieder zur Schule zurückkomme.
55. Ich träume nachts manchmal, daß andere Kinder meiner Klasse Dinge können, die ich nicht kann.
56. Wenn ich zu Hause daran denke, daß ich am nächsten Tag im Unterricht vorlesen muß, bekomme ich oft schon Angst, daß ich es schlecht machen könnte.
57. Wenn im Unterricht geprüft wird, spüre ich meistens ein komisches Gefühl im Magen.
58. Wenn ich etwas sehr schlecht gemacht habe, möchte ich am liebsten weinen, auch wenn ich versuche, mich zu beherrschen.
59. Manchmal träume ich, daß der Lehrer ärgerlich auf mich ist, weil ich meine Aufgaben nicht kann.
60. Ich habe Angst vor Klassenarbeiten.
61. Vor einer Klassenarbeit bin ich meist aufgeregt.
62. Nach jeder Klassenarbeit mache ich mir immer Gedanken darüber, ob ich wohl diesmal gut abgeschnitten habe.
63. Hinterher träume ich manchmal, daß ich bei der Klassenarbeit am Tag zuvor viele Fehler gemacht habe.
64. Bei Klassenarbeiten zittert mir oft die Hand.
65. Während einer Klassenarbeit mache ich oft Fehler, weil ich so aufgeregt bin.
66. Wenn der Lehrer sagt, daß wieder eine Klassenarbeit geschrieben wird, bekomme ich immer ein bißchen Angst.
67. Wenn eine schwere Klassenarbeit geschrieben wird, vergesse ich oft Dinge, die ich vorher gewußt habe.
68. Manchmal wünschte ich, daß ich mir nicht so viel Sorgen über Klassenarbeiten machte.
69. Wenn der Lehrer eine Klassenarbeit ankündigt, werde ich nervös.
70. Während einer Klassenarbeit denke ich gewöhnlich, daß ich es schlecht machen werde.
71. Wenn ich morgens zur Schule gehe, habe ich manchmal Angst, daß überraschend eine Klassenarbeit geschrieben wird.

5. MESSINSTRUMENT "SEMANTISCHES DIFFERENTIAL"

5.1. Allgemeine Hinweise durch den Versuchsleiter vor Beginn der Beantwortung der Fragen:

Die folgenden vier Fragen sind alle gleich - bis auf den ersten Satz in jeder Frage. Ihr habt die Aufgabe, Euch selbst und andere einzuschätzen, also zu sagen, welche Eigenschaften Ihr selbst und welche Eure Mitschüler haben.

Das ist ganz einfach, weil Ihr immer angeben müßt, welche Wörter am besten zu Euch passen. Ihr seht gleich auf dem Fragebogen eine lange Linie mit Zahlen dazwischen:

3---2---1---0---1---2---3

Rechts und links von dieser Linie stehen zwei Wörter. Es sind Gegensätze. Also zum Beispiel:

klein

groß

Ihr sollt nun zuerst sagen, welches dieser Wörter am besten zu Euch paßt. Paßt eher das Wort "groß" oder eher das Wort "klein"? Wenn Ihr Euch entschieden habt, welches der Wörter am besten zu Euch paßt, dann könnt Ihr noch angeben, wie gut das Wort paßt. Diese Angabe könnt Ihr machen, indem Ihr die Zahlen auf der langen Liste ankreuzt. Je näher die angekreuzte Zahl bei dem Wort steht, desto besser paßt es zu Euch.

Die Zahlen bedeuten im einzelnen:

3 = paßt sehr gut

2 = paßt ziemlich gut

1 = paßt etwas

0 = kann ich nicht sagen

Wenn also einer meint, er sei eher groß als klein, dann hat er sich schon erstmal für die Zahlen auf der rechten Seite von der "0" entschieden, die in der Nähe des Wortes "groß" stehen. Wenn er nun meint, das Wort "groß" passe sehr gut zu ihm, was kreuzt er dann wohl an? Und wenn er meint, das Wort "groß" passe nicht sehr gut, aber doch ziemlich gut zu ihm, was kreuzt er dann an? Was kreuzt aber einer an, der meint, das Wort "klein" passe "etwas" zu ihm?

Sobald die Antworten der Kinder zeigen, daß das Verfahren verstanden ist, wird der Fragebogen ausgeteilt. Die einzelnen Fragen werden noch einmal laut vor der Klasse verlesen. Die wesentlichen Schritte für die Bearbeitung werden noch einmal wiederholt. Nach Beginn der Bearbeitung werden nur Fragen der Kinder im persönlichen Gespräch beantwortet.

5.2. Schriftliche Instruktionen zu den Semantischen Differenzialen auf dem Schülerfragebogen

5.2.1. Konzept "SELBSTBILD"

Stelle Dir vor, Dein bester Freund oder Deine beste Freundin hat Dir erzählt, wie er (sie) so ist. Dein Freund (Deine Freundin) hat die guten, aber auch die schlechten Eigenschaften von sich geschildert.

Bitte sage Du nun ganz ehrlich und aufrichtig, wie Du selbst bist und welche Eigenschaften Du hast. Tue so, als ob Du es ganz allein zu Dir sagen würdest. (Wie Du weißt, erfährt kein anderer etwas von Deinen Antworten.)

Bitte kreuze in der folgenden Liste an, welche Wörter am besten zu Dir passen. Die Zahlen bedeuten:

- 3 = paßt sehr gut
- 2 = paßt ziemlich gut
- 1 = paßt etwas
- 0 = kann ich nicht sagen

5.2.2. Konzept "WUNSCHBILD"

Eben hast Du gesagt, wie Du jetzt bist. Bitte frage Dich jetzt: Bist Du zufrieden mit Dir? Oder bist Du vielleicht mit einigen Deiner Eigenschaften nicht so ganz zufrieden?

Sage nun bitte: Wie müßten Deine Eigenschaften sein, wenn Du mit Dir zufrieden sein willst?

Bitte kreuze in der folgenden Liste an, welche Wörter am besten zu Dir passen. Die Zahlen bedeuten:

- 3 = paßt sehr gut
- 2 = paßt ziemlich gut
- 1 = paßt etwas
- 0 = kann ich nicht sagen

5.2.3. Konzept "SCHWACHER SCHÜLER"

Diese Frage ist vielleicht ein bißchen schwer, aber Du kannst sie sicher beantworten.

Die Frage heißt: Wie sind die meisten Schüler, die in Englisch schwache Leistungen haben?

Versuche bitte, ihre Eigenschaften auf der Liste anzukreuzen!

Die Zahlen bedeuten:

- 3 = paßt sehr gut
- 2 = paßt ziemlich gut
- 1 = paßt etwas
- 0 = kann ich nicht sagen

5.2.4. Konzept "GUTER SCHÜLER"

Diese Frage kannst Du jetzt sicher auch beantworten!
Die Frage heißt: Wie sind die meisten Schüler, die in Englisch
gute Leistungen haben?

Versuche bitte, ihre Eigenschaften auf der Liste anzukreuzen!

Die Zahlen bedeuten:

- 3 = paßt sehr gut
- 2 = paßt ziemlich gut
- 1 = paßt etwas
- 0 = kann ich nicht sagen

5.2. Skalen, die einheitlich den Konzepten "SELBSTBILD", "WUNSCHBILD",
"SCHWACHER SCHÜLER", "GUTER SCHÜLER", "ENGLISCHLEHRER", "KLASSEN-
ARBEIT", "ZENSUR" und "KLASSENTTEST" zugrunde gelegt worden sind

fröhlich	3---2---1---0---1---2---3	traurig
weich	3---2---1---0---1---2---3	hart
freundlich	3---2---1---0---1---2---3	unfreundlich
gemächlich	3---2---1---0---1---2---3	eilig
ehrllich	3---2---1---0---1---2---3	unehrllich
nachgiebig	3---2---1---0---1---2---3	stark
gut	3---2---1---0---1---2---3	schlecht
ruhig	3---2---1---0---1---2---3	lebhaft
klug	3---2---1---0---1---2---3	dumm
zart	3---2---1---0---1---2---3	kraftvoll
angenehm	3---2---1---0---1---2---3	unangenehm
still	3---2---1---0---1---2---3	rege

6. Verzeichnis der Tabellen

1 : Aufteilung der Versuchsgruppen	50
2 : Ängstlichkeitsgrad und experimentelle Bedingung Klassentest vs. Klassenarbeit im 5. Schuljahr	51
3 : Ängstlichkeitsgrad und experimentelle Bedingung Klassentest vs. Klassenarbeit im 6. Schuljahr	51
4 : Homogenitätsprüfung der Stichprobenvarianzen bei den Variablen Intelligenz (LPS), Fremdspracheneignung (FTU) und Alter der Versuchsgruppen der Klassen 5 und 6	52
5 : Überprüfung der Nullhypothese der Stichprobenmittel- werte der je zwei Versuchsgruppen aus dem 5. und 6. Schuljahr bei den Variablen Intelligenz (LPS), Fremd- spracheneignung (FTU) und Alter	53
6 : F - Matrix	65
7 : P - Matrix (geordnet)	65
8 : Z - Matrix (geordnet)	66
9 : z - Differenzen - Matrix	67
10 : F - Matrix	68
11 : P - Matrix (geordnet)	69
12 : Z - Matrix (geordnet)	70
13 : Rangfolge der Behauptungen: (a) nach dem Schüler- und (b) nach dem Studentenurteil	72
14 : Koeffizienten der Aufgabenanalyse der TASC nach NICKEL, SCHLÜTER und FENNER (1973)	79
15 : Koeffizienten der Aufgabenanalyse der CMAS nach Schlüter und FENNER (1973)	80
16 : Mittelwerte und Standardabweichungen der Dimensionen Bewertung (B), Stärke (S) und Aktivität (A) der Kon- zepte IX - XVI (Selbstbild, Wunschbild, schwacher Schüler, guter Schüler, Englischlehrer, Klassenarbeit, Zensur und Klassentest) der Versuchsgruppe 5. Schul- jahr unter den experimentellen Bedingungen 'Klassen- arbeit' (KA-5) und 'Klassentest' (KT-5)	102
17 : Wie Tabelle 16, doch für die Versuchsgruppe 6. Sch.	103
18 : Variabilitätskoeffizienten: $V = 100 / \bar{x}$	104
19 : Nach den Variabilitätskoeffizienten berechnete Kor- relationen der experimentellen Bedingungen 'Klassen- arbeit' und 'Klassentest'	104
20 : Prüfung der Stichprobenvarianzen auf Homogenität: $F = s_1^2 / s_2^2$, geschätzte Streuungen der Grundgesamt- heit und t - Werte der Gruppen KA-5/KA-6, KA-5/ KT-6, KT-5/KA-6 und KT-5/KA-6 der Konzepte IX - XVI	106

21 : Konzept IX: t-Werte der Dimensionen Bewertung, Stärke und Aktivität der Gruppen KA-5/KA-6, KA-5/KT-6, KT-5/KA-6 und KT-5/KT-6	110
22 : Konzept X, s. Tab. 21	110
23 : Konzept XI, s. Tab. 21	112
24 : Konzept XII, s. Tab. 21	113
25 : Konzept XIII, s. Tab. 21	114
26 : Konzept XIV, s. Tab. 21	115
27 : Konzept XV, s. Tab. 21	116
28 : Konzept XVI, s. Tab. 21	116
29 : Tafel der Varianzanalyse: Konzept IX/ B-5	118
30 : Tafel der Varianzanalyse: Konzept IX/ S-5	118
31 : Tafel der Varianzanalyse: Konzept IX/ A-5	119
32 : Tafel der Varianzanalyse: Konzept X/B-5	119
33 : Tafel der Varianzanalyse: Konzept X/S-5	120
34 : Tafel der Varianzanalyse: Konzept X/A-5	121
35 : Tafel der Varianzanalyse: Konzept XI/B-5	122
36 : Tafel der Varianzanalyse: Konzept XI/S-5	122
37 : Tafel der Varianzanalyse: Konzept XI/A-5	122
38 : Tafel der Varianzanalyse: Konzept XII/B-5	123
39 : Tafel der Varianzanalyse: Konzept XII/S-5	124
40 : Tafel der Varianzanalyse: Konzept XII/A-5	124
41 : Tafel der Varianzanalyse: Konzept XIII/B-5	125
42 : Tafel der Varianzanalyse: Konzept XIII/S-5	125
43 : Tafel der Varianzanalyse: Konzept XIII/A-5	126
44 : Tafel der Varianzanalyse: Konzept XIV/B-5	127
45 : Tafel der Varianzanalyse: Konzept XIV/S-5	127
46 : Tafel der Varianzanalyse: Konzept XIV/A-5	127
47 : Tafel der Varianzanalyse: Konzept XV/B-5	128
48 : Tafel der Varianzanalyse: Konzept XV/S-5	129
49 : Tafel der Varianzanalyse: Konzept XV/A-5	129
50 : Tafel der Varianzanalyse: Konzept XVI/B-5	130
51 : Tafel der Varianzanalyse: Konzept XVI/S-5	130
52 : Tafel der Varianzanalyse: Konzept XVI/A-5	130
53 : Mittelwerte der Konzepte IX - XVI auf den Dimensionen Bewertung, Stärke und Aktivität	133
54 : Tafel der Varianzanalyse: Konzept IX/B-6	136
55 : Tafel der Varianzanalyse: Konzept IX/S-6	138
56 : Tafel der Varianzanalyse: Konzept IX/A-6	138

57 : Tafel der Varianzanalyse: Konzept X/B-6	139
58 : Tafel der Varianzanalyse: Konzept X/S-6	139
59 : Tafel der Varianzanalyse: Konzept X/A-6	139
60 : Tafel der Varianzanalyse: Konzept XI/B-6	140
61 : Tafel der Varianzanalyse: Konzept XI/S-6	141
62 : Tafel der Varianzanalyse: Konzept XI/A-6	141
63 : Tafel der Varianzanalyse: Konzept XII/B-6	142
64 : Tafel der Varianzanalyse: Konzept XII/S-6	142
65 : Tafel der Varianzanalyse: Konzept XII/A-6	143
66 : Tafel der Varianzanalyse: Konzept XIII/B-6	143
67 : Tafel der Varianzanalyse: Konzept XIII/S-6	144
68 : Tafel der Varianzanalyse: Konzept XIII/A-6	144
69 : Tafel der Varianzanalyse: Konzept XIV/B-6	145
70 : Tafel der Varianzanalyse: Konzept XIV/S-6	145
71 : Tafel der Varianzanalyse: Konzept XIV/A-6	146
72 : Tafel der Varianzanalyse: Konzept XV/B-6	147
73 : Tafel der Varianzanalyse: Konzept XV/S-6	147
74 : Tafel der Varianzanalyse: Konzept XV/A-6	148
75 : Tafel der Varianzanalyse: Konzept XVI/B-6	148
76 : Tafel der Varianzanalyse: Konzept XVI/S-6	149
77 : Tafel der Varianzanalyse: Konzept XVI/A-6	149
78 : Tafel der Varianzanalyse: Einstellung gegenüber England, den Engländern und der englischen Sprache im 5. Schuljahr	154
79 : Tafel der Varianzanalyse: Einstellung gegenüber England, den Engländern und der englischen Sprache im 6. Schuljahr	154

7. Verzeichnis der Abbildungen

1 : Der Bedingungs-zusammenhang von Beurteilungs- dimensionen	8
2 : Das Validitätsproblem	14
3 : Das Kontrollgruppenexperiment	28
4 : Formale Klassifizierung von Hypothesen	36
5 : Darstellung des Gehalts einer Hypothese in kom- plexen Sätzen	38
6 : Zweifaktorieller Versuchsplan der Untersuchung	45
7 : Habit-Interferenz-Theorie der Prüfungsangst von G. MANDLER und S.B. SARASON	75

8. Verzeichnis der Grafiken

1 : TASC/MASC	51
2 : Arithmetische Mittel der Dimensionen Bewertung, Stärke und Aktivität der Konzepte IX - XVI unter den experimentellen Bedingungen 'Klassenarbeit' und 'Klassentest' im 5. und 6. Schuljahr	109
3 : Arithmetische Mittel der Dimensionen Bewertung, Stärke und Aktivität der Konzepte IX - XVI unter den experimentellen Bedingungen 'Klassenarbeit' und 'Klassentest' bei wenig ängstlichen und sehr ängstlichen Schülern des 5. Schuljahrs	117
4 : Arithmetische Mittel der Dimensionen Bewertung, Stärke und Aktivität der Konzepte IX - XVI unter den experimentellen Bedingungen 'Klassenarbeit' und 'Klassentest' bei wenig ängstlichen und sehr ängstlichen Schülern des 6. Schuljahrs	137

LITERATURVERZEICHNIS

- ACHTENHAGEN, F.: Didaktik des fremdsprachlichen Unterrichts, Weinheim 1969.
- ACHTENHAGEN, F.: Möglichkeiten und Grenzen einer Taxonomie von Lernzielen, in: Die Deutsche Berufs- und Fachschule, 1969, S. 693 - 706.
- ADORNO, Th. W. u.a.: Der Positivismusstreit in der deutschen Soziologie, Neuwied und Berlin 1972².
- BARRUTIA, R.: Die Linguistik und ihr methodisches Risiko, in: Freudenstein, R./Gutschow, H. (Hrsg.): Fremdsprachen, Lehren und Erlernen, München 1972.
- BARTMANN, Th.: Der Einfluß von Zeitdruck auf die Leistung und das Denkverhalten von Volksschülern, in: Psychol. Forschung, 1963.
- BELSER, H.: Testentwicklung, Weinheim 1967.
- BENNER, D.: Hauptströmungen der Erziehungswissenschaft. Eine Systematik traditioneller und moderner Theorien, München 1973.
- BERGER, H.: Erfahrung und Gesellschaftsform. Methodologische Probleme wissenschaftlicher Beobachtung, Stuttgart/Berlin/Köln/Mainz 1972.
- BERNSTEIN, B.: Soziokulturelle Determinanten des Lernens, in: Heintz, P. (Hrsg.): Soziologie der Schule, Köln/Opladen 1959.
- BIGLMAIER, F.: Leistungsmessung durch informelle Lebertests, in: betrifft: Erziehung, 1969, S. 22 - 26 und S. 26 - 28.
- BLANKERTZ, H.: Pädagogische Theorie und empirische Forschung, in: Zur Bedeutung der Empirie für die Pädagogik als Wissenschaft. In: Neue Folge der Ergänzungshefte zur Vierteljahresschrift für wissenschaftliche Pädagogik, Heft 5, 1966, S. 65 - 78.
- BLANKERTZ, H.: Theorien und Modelle der Didaktik, München 1969².
- BLANKERTZ, H.: Pädagogik unter wissenschaftstheoretischer Kritik, in: Oppolzer, S. (Hrsg.): Erziehungswissenschaft 1971 zwischen Herkunft und Zukunft der Gesellschaft, Wuppertal 1971, S. 20 - 33.
- BLOOM, B. S. (ed.): Taxonomy of Educational Objectives, Handbook I: Cognitive Domain, New York 1956.
- BREDENKAMP, J.: Experiment und Feldexperiment, in: Graumann, C. F. (Hrsg.): Handbuch der Psychologie, Band 7/1. Halbband, Sozialpsychologie, Göttingen 1969, S. 332 - 374.
- BREDENKAMP, J.: Der Signifikanztest in der psychologischen Forschung, Frankfurt am Main 1972.
- BRENTANO, M. von: Die unbescheidene Philosophie, in: Argumente 43, 1967, S. 102 - 116.
- BRIM, O. G.: Soziologie des Erziehungswesens. Pädagogische Forschungen. Veröffentlichungen des Comenius-Instituts, Band 24, Heidelberg 1963.
- BÜNTING, K.-D.: Einführung in die Linguistik, Frankfurt 1971.

- BUHR, M./KLAUS, G.: Philosophisches Wörterbuch, Berlin 1972⁸.
- CASTANEDA, A./McCANDLESS, B.R./PALERMO, D.S.: The children's form of manifest anxiety scale, in: Child Developm., 1956, S. 317-326.
- CHAUNCEY, H./DOBBIN, J.E.: Der Test im modernen Bildungswesen, Stuttgart 1968.
- CICOUREL, A.V.: Methode und Messung in der Soziologie, Frankfurt am Main 1970.
- CLAUSS, G./EBNER, H.: Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen, Berlin 1968.
- DAVIES, A.: Language Testing Symposium. A Psycholinguistic Approach, London 1968.
- DENNINGHAUS, F.: Die wechselseitigen Einflüsse zwischen der Linguistik und dem Fremdsprachenunterricht, in: Praxis des neusprachlichen Unterrichts, 1971, S. 31 - 40.
- DICK, F.: wenn p, dann q. die widersprüche des nomothetischen gesetzsbegriffs der empirisch-analytischen sozialwissenschaften und dessen doppelt ideologische funktion (zum theorienstreit zwischen popper & holzkamp + zur realität der forschungspraxis). theorie + kritik 1, Giessen 1972.
- DIEDERICH, P.B.: Short - cut Statistics for Teacher-made Tests. Educational Testing Service, Evaluation and Advisory Service Series. No. 5, 1964²; deutsch: Diederich, P.B.: Statistische Kurzverfahren zur Analyse informeller Tests, in: Chauncey, H./Dobbin, J.E.: Der Test im modernen Bildungswesen, Stuttgart 1968.
- DIEDERICH, J.: Hilfen zur Kritik empirischer Untersuchungen, in: Zeitschrift für Pädagogik, 1970, S. 247 - 264.
- DIETRICH, G.: Bildungswirkungen des Gruppenunterrichts. Persönlichkeitsformende Bedeutung des gruppenunterrichtlichen Verfahrens, München 1969.
- DOHSE, W.: Das Schulzeugnis. Sein Wesen und seine Problematik, Weinheim 1963.
- DORSCH, F.: Psychologisches Wörterbuch, Hamburg/Bern 1963⁷.
- DREITZEL, H.P.: Soziologische Reflexionen über das Elend des Leistungsprinzips, in: Sinn und Unsinn des Leistungsprinzips. Ein Symposium, München 1974.
- DRENTH, P.: Der psychologische Test, München 1969.
- ECKEL, K.: Das pädagogische Methodenexperiment, in: Programmierteres Lernen und Programmierter Unterricht, 1970, S. 156 - 173.
- ECKEL, K.: Die Bedeutung des Klasseneffekts für die schulpädagogische Forschung, in: Programmierteres Lernen und programmierter Unterricht, 1969, S. 97 - 113.
- EDWARDS, A.L.: Techniques of Attitude Scale Construction, New York 1957.
- EKMAN, G.: Konstruktion und Standardisierung von Tests, Göttingen 1955.

- ERTEL, S.: Die emotionale Natur des semantischen Raumes, in: Psychologische Forschung 1964.
- ESSER, U.: Skalierungsverfahren, in: Methoden der marxistisch-leninistischen Sozialforschung, herausgegeben von: Friedrich, F., Berlin 1971, S. 184 - 273.
- ETHERTON, A.R.B.: Objective English Tests for Secondary Schools, London 1969⁴.
- FIJALKOWSKI, J.: Über einige Theorie-Begriffe in der deutschen Soziologie der Gegenwart, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 1961, S. 88 - 109.
- FISCHER, G. (Hrsg.): Psychologische Testtheorie, Bern/Stuttgart 1968.
- FLECHSIG, K.-H.: Forschung im Bereich des Fremdsprachenunterrichts, in: Ingenkamp, K./Parey, E. (Hrsg.): Handbuch der Unterrichtsforschung, Teil III, Weinheim/Berlin/Basel 1971.
- FRICKE, R.: Zur Theorie des Rasch-Modells und einige Erfahrungen bei der Anwendung dieses Modells auf Schulleistungstests, in: ZeF 1972, S. 3 - 23.
- FRICKE, R.: Über Meßmodelle in der Schulleistungsdiagnostik, Reihe: Studien zur Lehrforschung, Band 2, Düsseldorf 1972.
- FRÜCHTER, B.: Introduction to Factor Analysis, New Jersey 1968⁸.
- FUCHS, A.: Das Eindrucksdifferential als Instrument zur Erfassung emotionaler Bedeutungsprozesse, in: Bergler, R. (Hrsg.): Beiträge zur empirischen Sozialforschung. Das Eindrucksdifferential. Theorie und Technik, 1975, S. 69 - 100.
- FÜRSTENAU, P.: Zur Psychoanalyse der Schule als Institution, in: Pädagogisches Zentrum. Veröffentlichungen. Hrsg.: C.-L. Furck. Reihe B: Diskussionsbeiträge. Bd. 10. Zur Theorie der Schule. Mit Beiträgen von P. Fürstenau u.a., Weinheim/Berlin/Basel 1969.
- FÜRSTENAU, P.: Neuere Entwicklungen der Bürokratieforschung und das Schulwesen. Ein organisationssoziologischer Beitrag, in: Pädagogisches Zentrum. Veröffentlichungen. Hrsg.: C.-L. Furck. Reihe B: Diskussionsbeiträge. Bd. 10. Zur Theorie der Schule. Mit Beiträgen von P. Fürstenau u.a., Weinheim/Berlin/Basel 1969.
- GÄRTNER-HARNACH, V.: Angst und Leistung, Weinheim 1972.
- GAMM, H.-J.: Kritische Schule. Eine Streitschrift für die Emanzipation von Lehrern und Schülern, München 1970.
- GAUDE, P./TESCHNER, W.: Objektivierter Leistungsmessung in der Schule. Einsatz informeller Tests im leistungsdifferenzierten Unterricht, Frankfurt/Berlin/München 1970.
- GEER, J.P. van de: Introduction to Multivariate Analysis for the Social Sciences, San Francisco 1971.

- GEHLERT, S./FREYN, R.: LPS-Profile 14-jähriger Schüler, in: Dietrich, G. (Hrsg.): Unterrichtspsychologie der Sekundarstufe. Beiträge zu Forschung, Theorie und Praxis, Donauwörth 1972.
- GLASER, R.: Unterrichtstechnologie und die Messung von Lernergebnissen, in: Issing, L.J.I. (Hrsg.): Der programmierte Unterricht in den USA heute. Weinheim/Berlin 1967, S. 43 - 50.
- GLINZ, H.: Linguistische Grundbegriffe und Methodenüberblick, Frankfurt/M. 1971.
- GÜLLER, A.: Zensuren und Zeugnisse, Stuttgart 1966.
- GÖTZ, D./BURGSCHMIDT, E.: Einführung in die Sprachwissenschaft für Anglisten, München 1971.
- GRONLUND, N.E.: Measurement and Evaluation in Teaching, New York 1965.
- GUILFORD, J.P.: Psychometric Methods, New York 1954².
- GUILFORD, J.P.: Fundamental Statistics in Psychology and Education, New York/Sydney 1964⁴.
- GUSKIN, A.E./Guskin, S.L.: Sozialpsychologie in Schule und Unterricht, Ulm 1973.
- GUTJAHR, W.: Die Messung psychischer Eigenschaften, Berlin 1961.
- GUTSCHOW, H.: Englisch an Volksschulen. Probleme und Arbeitsformen, Berlin/Bielefeld 1964.
- HABERMAS, J.: Analytische Wissenschaftstheorie und Dialektik. Ein Nachtrag zur Kontroverse zwischen Popper und Adorno, in: Topitsch, E. (Hrsg.): Logik der Sozialwissenschaften, Köln/Berlin 1968⁵, S. 291 - 311.
- HABERMAS, J.: Zur Logik der Sozialwissenschaften, Frankfurt am Main 1970.
- HEIMANN, P./OTTO, G./SCHULZ, W.: Unterricht - Analyse und Planung, Hannover 1965.
- HEINZE, Th.: Zur Kritik an den Technologisierungstendenzen des Unterrichtsprozesses, in: Die Deutsche Schule, 1972, S. 347 - 361.
- HELM, J.: Über den Einfluß affektiver Spannungen auf das Denken, in: Zeitschr. f. Ps., 1954.
- HERRMANN, Th.: Persönlichkeitsmerkmale. Bestimmung und Verwendung in der psychologischen Wissenschaft, Stuttgart/Berlin/Köln/Mainz 1973.
- HEUER, H.: England in Vorstellung und Vor-Wissen unserer Schüler. Die Ergebnisse einer empirischen Untersuchung, in: Englisch, 1966.
- HEUER, H.: Zur Motivation im Englischunterricht. Die Ergebnisse einer empirischen Untersuchung, in: Englisch, 1968.
- HÖRMANN, H.: Psychologie der Sprache, Berlin/Heidelberg/New York 1967.

- HORST, P.: Messung und Vorhersage. Eine Einführung in die psychologische Testtheorie, Weinheim/Berlin/Basel 1971.
- HOFSTÄTTER, P.R.: Tiefenpsychologische Persönlichkeitstheorien, in: Lersch, Ph./Thomae, H. (Hrsg.): Handbuch der Psychologie, Bd. 4, Persönlichkeitsforschung und Persönlichkeitstheorie, Göttingen 1960², S. 385 - 414.
- HOLM, K.: Zuverlässigkeit von Skalen und Indizes, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 1970, S. 356 - 386 und S. 693 - 714.
- HOLZKAMP, K.: Wissenschaftstheoretische Voraussetzungen kritisch-emanzipatorischer Psychologie, in: Zeitschrift für Sozialpsychologie, 1970, S. 5 - 21 und S. 109 - 141.
- HOLZKAMP, K.: Zum Problem der Relevanz psychologischer Forschung für die Praxis, in: Psychologische Rundschau, 1970, S. 1 - 22.
- HÜLLEN, W.: Linguistik und Englischunterricht. Didaktische Analysen, Heidelberg 1971.
- HURRELMANN, K.: Unterrichtsorganisation und schulische Organisation. Eine empirische Untersuchung zur Rolle der "Leistungsdifferenzierung" im schulischen Selektionsprozeß, Weinheim/Berlin/Basel 1971.
- INGENKAMP, K.: Die deutschen Schulleistungstests, Weinheim 1962.
- INGENKAMP, K.: Möglichkeiten und Grenzen des Lehrerurteils und der Schultests, in: Roth, H. (Hrsg.): Begabung und Lernen, Gutachten und Studien der Bildungskommission des Deutschen Bildungsrates, Stuttgart 1968, S. 407 - 431.
- INGENKAMP, K.: Die Fragwürdigkeit der Zensurengebung, Weinheim 1971².
- INGRAM, E.: Appendix. Item Analysis, in: Davies, A. (ed.): Language Testing Symposium. A Psycholinguistic Approach, London 1968, S. 192 - 203.
- ITELSON, L.: Mathematische und kybernetische Methoden in der Pädagogik, Berlin 1967.
- KAMRATOWSKI, J./PENNE, K.J./SCHNEIDER, J.: Informelle Fremdsprachentests für die Schulpraxis, Berlin 1970.
- KERLINGER, F.N.: Foundations of Behavioral Research, New York 1970³.
- KLAUER, K.J.: Das Experiment in der pädagogischen Forschung, Düsseldorf 1973.
- KLAUS, G.: Moderne Logik. Abriß der formalen Logik, Berlin 1970⁵.
- KLÜVER, J./KRÜGER, H.: Aktionsforschung und soziologische Theorien. Wissenschaftstheoretische Überlegungen zum Erkenntnisinteresse in der Aktionsforschung, in: Haag, F. u.a.: Aktionsforschung. Forschungsstrategien, Forschungsfelder und Forschungspläne, München 1972, S. 76 - 99.
- KRECH, D./CRUTCHFIELD, R.S.: Grundlagen der Psychologie, Band I, Weinheim/Berlin/Basel 1969².

- KRECKEL, R.: Soziologische Erkenntnis und Geschichte. Über Möglichkeiten und Grenzen einer empirisch-analytischen Orientierung in den Humanwissenschaften, Opladen 1972.
- KROPE, P.: Zensur, Ausbildung und Gesellschaft. Zur Kritik und Alternative eines Bewertungsverfahrens, Essen 1976.
- KRUMM, V.: Leistung - Schulleistung - Schulleistungsmessung, in: Wulf, Chr. (Hrsg.): Wörterbuch der Erziehung, München/Zürich 1974.
- KVALE, S.: Prüfung und Herrschaft. Hochschulprüfungen zwischen Ritual und Rationalisierung, Weinheim/Basel 1972.
- LADO, R.: Language Testing, London 1967.
- LEVITT, E.E.: Die Psychologie der Angst, Stuttgart/Berlin/Mainz/Köln 1971.
- LEWIN, K.: Field Theory in Social Science, New York 1951.
- LIENERT, G.A.: Testaufbau und Testanalyse, Weinheim 1967².
- LINDER, A.: Planen und Auswerten von Versuchen. Eine Einführung für Naturwissenschaftler, Mediziner und Ingenieure, Basel und Stuttgart 1959².
- LORD, F.M./NOVICK, M.R.: Statistical Theories of Mental Test Scores. With Contributions by A. Birnbaum, Massachusetts/London/Ontario 1968.
- LORGE, I.: The Fundamental Nature of Measurement, in: Lindquist, E.F. (ed.): Educational Measurement, Washinton 1963⁵.
- LYONS, J.: Einführung in die moderne Linguistik, München 1971.
- MAGNUSSON, D.: Testtheorie, Wien 1969.
- MAYNTZ, R./HOLM, K./HÜBNER, P.: Einführung in die Methoden der empirischen Soziologie, Köln und Opladen 1969.
- MAYNTZ, R./ZIEGLER, R.: Zur Soziologie der Organisation, in: König, R. (Hrsg.): Handbuch der empirischen Sozialforschung, Band II, Stuttgart 1969.
- MEILE, B.: Voraussetzungen der empirischen Pädagogik. Ein Wissenschaftstheoretischer Aufriß, in: Zeitschrift für Pädagogik, 1972, S. 709 - 733.
- MESSELKEN, H.: Empirische Sprachdidaktik, Heidelberg 1971.
- MESSNER, R.: Die Messung schulischen Erfolgs. Eine Einführung in die Probleme der Kontrolle des Unterrichtserfolgs durch den Lehrer, in: Messner, R./Rumpf, H. (Hrsg.): Didaktische Impulse. Studentexte zur Analyse von Unterricht, Wien 1971, S. 145 - 174.
- METZGER, W.: Das Experiment in der Psychologie, in: Studium Generale, 1952, S. 142 - 163.
- METZGER, W.: Stimmung und Leistung. Die affektiven Grundlagen des Lernerfolgs, Münster 1967⁴.
- MEYER, H.L.: Einführung in die Curriculum-Methodologie. Mit einem Vorwort von H. Blankertz, München 1972.

- MEYER, H.L.: Trainingsprogramm zur Lernzielanalyse, Frankfurt am Main 1974.
- MICHEL, L.: Allgemeine Grundlagen psychometrischer Tests, in: Heiss, R. (Hrsg.): Handbuch der Psychologie, Band 6, Psychologische Diagnostik, Göttingen 1964, S. 19 - 70.
- MOLLENHAUER, K.: Pädagogik und Rationalität, in: Mollenhauer, K.: Erziehung und Emanzipation, München 1968, S. 55 - 74.
- MOLLENHAUER, K.: Theorien zum Erziehungsprozeß. Zur Einführung in erziehungswissenschaftliche Fragestellungen, München 1972.
- MOLLENHAUER, K./RITTELMAYER, Ch.: "Empirisch-analytische Wissenschaft" versus "Pädagogische Handlungsforschung": eine irreführende Alternative, in: Zeitschr. f. Pädagogik, 1975, S. 69ff.
- MOSER, H.: Leistungsprinzip: Ökonomische Kategorie und soziale Norm, in: Die Deutsche Schule, 1972, S. 202 - 212.
- MÜHLE, G.: Furcht und Angst unter entwicklungspsychologischem Aspekt, in: Wiesbrock, H. (Hrsg.): Die politische und gesellschaftliche Rolle der Angst.
- MÜLLER, M.: Nachwort, in: Sinn und Unsinn des Leistungsprinzips. Ein Symposium, München 1974.
- MÜLLER, U.: Zur möglichen Bedeutung der empirischen Sozialforschung für die Marxsche Theorie, in: Backhaus, H.-G. u.a. (Hrsg.): Gesellschaft - Beiträge zur Marxschen Theorie 5, Frankfurt a. M. 1975.
- MÜNCH, W.: Datensammlung in den Sozialwissenschaften. Eine Einführung in ihre Methoden, Stuttgart/Berlin/Köln/Mainz 1971.
- NEANDER, J.: Objektivierter Lernerfolgsmessung in der Gesamtschule - Fortschritt für wen? In: Die Deutsche Schule, 1973, S. 35 - 47.
- NEUHAUS, E.: Zum pädagogischen Leistungsbegriff, in: Lichtenstein-Rother, I. (Hrsg.): Schulleistung und Leistungsschule, Bad Heilbrunn/Obb. 1971.
- NEUMANN, G.: Artikel "Arbeit", in: Wulf, Chr. (Hrsg.): Wörterbuch der Erziehung, München/Zürich 1974, S. 30 - 35.
- NICKEL, H./SCHLÜTER, P.: Angstwerte bei Hauptschülern und ihr Zusammenhang mit Leistungs- sowie Verhaltensmerkmalen, Lehrerurteil und Unterrichtstil, in: Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 1970, S. 125 - 136.
- NICKEL, H./SCHLÜTER, P./PENNER, H.-J.: Angstwerte, Intelligenztest- und Schulleistungen sowie der Einfluß der Lehrerpersönlichkeit bei Schülern verschiedener Schularten, in: Psychologie in Erziehung und Unterricht, 1973, S. 1 - 13.
- OERTER, R.: Struktur und Wandlung von Werthaltungen, München 1970.
- OERTER, R.: Moderne Entwicklungspsychologie, Donauwörth 1967.
- OFFE, C.: Leistungsprinzip und industrielle Arbeit, Fr. a. M. 1970.

- OEVERMANN, U.: Sprache und soziale Herkunft. Ein Beitrag zur Analyse schichtenspezifischer Sozialisationsprozesse und ihrer Bedeutung für den Schulerfolg, Frankfurt a.M. 1972.
- OEVERMANN, U.: Schichtenspezifische Formen des Sprachverhaltens und ihr Einfluß auf die kognitiven Prozesse, in: Roth, H. (Hrsg.): Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen. Deutscher Bildungsrat: Gutachten und Studien der Bildungskommission, Band 4, Stuttgart 1969⁴.
- OPP, K.-D.: Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theoriebildung, Hamburg 1970.
- ORLIK, P.: Kritische Untersuchungen zur Begabtenförderung, Düsseldorf 1965.
- ORLIK, P.: Eine Technik zur erwartungstreuen Skalierung psychologischer Merkmalsräume aufgrund von Polaritätsprofilen, in: Z. f. experim. angew. Ps., 1967.
- OSGOOD, Ch. E./SUCI, G.J./TANNENBAUM, P.H.: The Measurement of Meaning, Illinois 1971⁸.
- PETERSEN, H.: Reformbestrebungen im Englischunterricht und ihre Kriterien, in: Linguistik und Didaktik, 1970.
- PETERSEN, H.: Überlegungen zur Anwendung der generativen Transformationsgrammatik im Englischunterricht an deutschen Schulen, in: Der fremdsprachliche Unterricht, 1971, S. 25 - 48.
- PIMSLEUR, P.: Language Aptitude Testing, in: Davies, A. (ed.): Language Testing Symposium. A Psycholinguistic Approach, London 1968.
- PIMSLEUR, P.: Testing Foreign Language Learning, in: Valdman, A. (ed.): Trends in Language Teaching, New York 1966.
- PRELL, S.: Die Auswirkung von Prüfungsangst auf verschiedene Formen der Lernzielkontrolle, in: Psychologie in Erziehung und Unterricht, 1973, S. 14 - 30.
- RAUCH, E./ANZINGER, W.: Wörterbuch Kritische Erziehung, Starnberg 1972.
- ROEDER, P.-M.: Sprache, Sozialstatus und Bildungschancen, in: Roeder, P.-M./Pasdzierny, A./Wolf, W.: Sozialstatus und Schulerfolg, Heidelberg 1965, S. 5 - 32.
- ROEDER, P.-M.: Thesen zur Auslese durch den neusprachlichen Unterricht, in: Roeder, P.-M. (Hrsg.): Pädagogische Analysen und Reflexionen. Festschrift für Elisabeth Blochmann zum 75. Geburtstag, Weinheim und Berlin 1967, S. 309 - 327.
- ROEDER, P.-M.: Die pädagogische Bedeutung von Erfolgskontrolle und Leistungsmessung, in: Speck, J. (Hrsg.): Leistung, Erfolg und Erfolgskontrolle in der Pädagogik und ihren Nachbarwissenschaften, Münster 1968.
- ROTH, E.: Einstellung als Determination individuellen Verhaltens, Göttingen 1967.
- ROTH, E.: Persönlichkeitspsychologie. Eine Einführung. Stuttgart/Berlin/Köln/Mainz 1969.

- ROTH, H.: Begabung und Lernen. Ergebnisse und Folgerungen neuer Forschungen. Deutscher Bildungsrat: Gutachten und Studien der Bildungskommission, Band 4, Stuttgart 1969⁴.
- SACHS, L.: Statistische Auswertungsmethoden, 1969².
- SALZMANN, Chr.: Studien zu einer Theorie des Prüfens und Erprobens im Raum der Erziehung, Ratingen 1967.
- SAMSTAG, K./ BAUS, M.: Pädagogisch-psychologisches Testen, Bad Heilbrunn 1962.
- SARASON, S.B./DAVIDSON, K.S./LIGHTHALL, F.F./WAITE, R.R./RUEBUSH, B.K.: Angst bei Schulkindern. Ein Forschungsbericht, Stuttgart 1971.
- SECORD, P.F./BACKMAN, C.W.: Social Psychology, New York 1964.
- SEGETH, W.: Elementare Logik, Berlin 1972⁷.
- SEIBEL, H.D.: Gesellschaft im Leistungskonflikt, Düsseldorf 1973.
- SEILER, H.: Moderne Forschungstheorie und Erziehungswissenschaft. Bemerkungen zum Programm der empirischen Pädagogik, in: Zeitschrift für Pädagogik, 1969, S. 501 - 523.
- SIEGEL, S.: Nonparametric Statistics, New York 1956.
- SINGER, K.: Lernhemmung, Psychoanalyse und Schulpädagogik, München 1970.
- SIXTL, F.: Meßmethoden der Psychologie, Weinheim 1967.
- SIXTL, F.: Beiträge zur Explikation und Kritik der klassischen Testtheorie, in: Fischer, C.W. (Hrsg.): Psychologische Testtheorie, Wien/Bern/Stuttgart 1963, S. 191 - 212.
- SIXTL, F.: Kritik an der theoretischen Konzeption der Leistungstests: Die faktische Inkonsistenz der Testtheorie, in: Ingenkamp, K./Marsolek, Th.: Möglichkeiten und Grenzen der Testanwendung in der Schule, Weinheim/Berlin/Basel 1968, S. 467 - 475.
- SÜLLWOLD, F.: Schultests, in: Heiss, R. (Hrsg.): Handbuch der Psychologie, Band 6, Diagnostik, Göttingen 1964.
- SÜLLWOLD, F.: Theorie und Methodik der Einstellungsmessung, in: Graumann, C.F. (Hrsg.): Handbuch der Psychologie, Band 7/1. Halbband, Sozialpsychologie, Göttingen 1969, S. 475 - 514.
- SUPPES, P./ZINNES, J.L.: Basic Measurement Theory, in: Luce, R.D./Bush, R.R./Galanter, E. (eds.): Handbook of Mathematical Psychology, Vol. 1, New York/London 1963, S. 3 - 76.
- SCHELL, H.: Angst und Schulleistung, Göttingen 1972.
- SCHELSKY, H.: Soziologische Bemerkungen zur Rolle der Schule in unserer Gesellschaftsverfassung. Eine Denkschrift (1956), in: Auf der Suche nach der Wirklichkeit, Düsseldorf/Köln 1965, S. 131 - 159.

- SCHEUCH, E.K.: Skalierungsverfahren in der Sozialforschung, in: König, R. (Hrsg.): Handbuch der empirischen Sozialforschung, Band I, Stuttgart 1967², S. 348 - 384.
- SCHREINER, G.: Sinn und Unsinn der schulischen Leistungsbeurteilung, in: Die Deutsche Schule, 1971, S. 226 - 237.
- SCHULZ, W.: Kausalität und Experiment in den Sozialwissenschaften, Methodologie und Forschungstechnik, München 1970.
- SCHWARZ, E.: Experimentelle und quasi-experimentelle Anordnungen in der Unterrichtsforschung, in: Ingenkamp, K./Parey, E. (Hrsg.): Handbuch der Unterrichtsforschung, Teil I, Weinheim 1970.
- STEGMÜLLER, W.: Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Studienausgabe, Teil I/III und V, Berlin/Heidelberg/New York 1969.
- STOSBERG, M.: Analyse der Massenkommunikation: Einstellungsmessung, Düsseldorf 1972.
- TENT, L.: Die Auslese von Schülern für weiterführende Schulen, Göttingen 1969.
- THIEMANN, F.: Der Beitrag empirischer Unterrichtsforschung für die Konzeption von Unterricht, Bad Heilbrunn 1973.
- TRAVERS, R.M.W.: Einführung in die erziehungswissenschaftliche Forschung, München 1972.
- TORGERSON, W.S.: Theory and Methods of Scaling, New York/London/Sydney 1967⁷.
- ULICH, D.: Gruppendynamik in der Schulklasse. Möglichkeiten und Grenzen sozialwissenschaftlicher Analysen, München 1971.
- ULICH, D./MERTENS, W.: Urteile über Schüler. Zur Sozialpsychologie pädagogischer Diagnostik, Weinheim/Basel 1973.
- ÜBERLA, K.: Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler, Berlin/Heidelberg/New York 1971².
- VALETTE, R.M.: Tests im Fremdsprachenunterricht, Berlin/Bielefeld 1971.
- VUKOVICH, A.F.: Die Konstruktion psychologischer Tests, in: Heiss, R. (Hrsg.): Handbuch der Psychologie, Band 6, Diagnostik, Göttingen 1964, S. 113 - 144.
- WEINERT, F.E.: Schule und Beruf als institutionelle Sozialisationsbedingungen, in: Graumann, C.F. (Hrsg.): Handbuch der Psychologie, Band 7/2. Halbband, Sozialpsychologie, Göttingen 1972, S. 825 - 885.
- WEINGARTNER, P.: Wissenschaftstheorie I. Einführung in die Hauptprobleme, Bad Cannstatt 1971.
- WEIS, V.: Zensierungsmodelle und ihre pädagogischen Konsequenzen, in: Die Deutsche Schule, 1971, S. 542 - 553.

- WELLENDORF, F.: Schulische Sozialisation und Identität. Zur Sozialpsychologie der Schule als Institution, Weinheim/Basel 1974².
- WELLENDORF, F.: Ansätze zur erziehungswissenschaftlichen Theoriebildung in der Bundesrepublik, in: Furck, C.-L./Goldschmidt, D./Röbbelen, I. (Hrsg.): Gesellschaft und Erziehung I, Heidelberg 1969, S. 68 - 110.
- WENDELER, J.: Standardarbeiten - Verfahren zur Objektivierung der Notengebung, Weinheim/Berlin/Basel 1969.
- WERNER, H.: Einführung in die Entwicklungspsychologie, München 1959⁴.
- WESTMEYER, H.: Kritik der psychologischen Unvernunft. Probleme der Psychologie als Wissenschaft, Stuttgart/Berlin/Köln/Mainz 1973.
- WINNEFELD, F.: Pädagogischer Kontakt und pädagogisches Feld, München/Basel 1963.
- WITTE, W.: Das Problem der Bezugssysteme, in: Metzger, W. (Hrsg.): Handbuch der Psychologie, Band I/1. Halbband, Allgemeine Psychologie, Göttingen 1966, S. 1003 - 1027.

LEBENS LAUF

Ich wurde am 3. Dezember 1939 in Zossen/Teltow in Brandenburg geboren. Von 1946-1954 besuchte ich die Volksschulen in Wehrhain und Schlieben, Sachsen-Anhalt.

Im Jahre 1954 verließen meine Eltern die DDR. Während des Aufenthalts im Flüchtlingslager Lübeck-Blankensee nahm ich am Unterricht der Lagerschule teil.

Nach erfolgreich bestandener Aufnahmeprüfung besuchte ich von 1955-1961 das Staatliche Aufbaugymnasium Rüthen/Möhne. Dort erhielt ich am 13.3.1961 das Reifezeugnis.

Im Sommersemester 1961 immatrikulierte ich mich an der Pädagogischen Akademie Dortmund, an der ich am 13.3.1964 die Erste Prüfung für das Lehramt an Volksschulen bestand. Am 1.4.1964 trat ich als Lehrer z.A. in Bielefeld in den Schuldienst ein. Im Dezember 1965 legte ich eine Zusatzprüfung ab für das Fach Englisch an Volksschulen. Die Zweite Lehrerprüfung bestand ich am 5.12.1967. Die Ernennung zum Volksschullehrer und Beamten auf Lebenszeit erfolgte am 19.4. 1968.

Im Sommersemester 1968 immatrikulierte ich mich an der Westfälischen Wilhelms - Universität Münster. Ich begann das Studium der Psychologie. Ab Wintersemester 1968/69 studierte ich die Fächer Pädagogik und Soziologie.

Vom 25.7.1969 - 31.12.1971 wurde ich für einen Einsatz im Rahmen eines Forschungsvorhabens zur Verbesserung des Englischunterrichts an weiterführenden Schulen an die Pädagogische Hochschule Westfalen - Lippe, Abt. Bielefeld, abgeordnet. Darauf folgte die Abordnung als Förderassistent vom 1.1.1972 - 31. 12.1974.

Vom 1.6.1970 an übte ich eine einjährige Tätigkeit als Fachleiter für Deutsch am Bezirksseminar Brackwede/Bielefeld aus.

Am 1.1.1975 nahm ich den Dienst als Volksschullehrer wieder auf.

Seit dem 13. September 1976 bin ich Schulleiter der evang. Grundschule Warstein 2 - Belecke.

Meine Dissertation mit dem Thema 'Affektive Wirkungen norm-orientierter Prüfinstrumente im Fach Englisch auf die Einstellung sehr ängstlicher und wenig ängstlicher Schüler.

Eine experimentell angelegte Pilotstudie im 5. und 6. Schuljahr' habe ich unter der Betreuung von Herrn Prof. Dr. H. Blankertz angefertigt.

